



HAL
open science

Qualité et valeur des prévisions solaires probabilistes

Josselin Le Gal La Salle

► **To cite this version:**

Josselin Le Gal La Salle. Qualité et valeur des prévisions solaires probabilistes. Physique Numérique [physics.comp-ph]. Université de la Réunion, 2021. Français. NNT : 2021LARE0016 . tel-03373288

HAL Id: tel-03373288

<https://theses.hal.science/tel-03373288>

Submitted on 11 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

École doctorale n° 542 : Sciences Technologies et Santé

Section 62 - Spécialité Énergétique

THÈSE DE DOCTORAT de l'Université de La Réunion

présentée et soutenue publiquement par

Josselin Le Gal La Salle

le 1^{er} juillet 2021

Qualité et valeur des prévisions solaires probabilistes

Jury

M. Philippe Lauret,	Professeur, Université de La Réunion	Directeur de thèse
M. Mathieu David,	Docteur HDR, Université de La Réunion	Co-encadrant
M. Richard Perez,	Professeur, University at Albany, State of New York	Rapporteur
M. Stéphane Grieu,	Professeur, Université de Perpignan	Rapporteur
M. Philippe Blanc,	Directeur de recherche, MINES ParisTech	Examineur
M. Pierre Pinson,	Professeur, Technical University of Denmark	Examineur
M. Jordi Badosa,	Ingénieur de recherche, LMD, École Polytechnique	Examineur

Résumé étendu

La prévision solaire est un élément central de la plupart des solutions d'intégration à large échelle de l'énergie solaire au sein des mix électriques. L'objectif de ce travail est de démontrer que l'utilisation de prévisions probabilistes, encore peu fréquente à l'échelle industrielle, est pertinente pour de nombreuses applications. Comme la littérature ayant trait à la valeur des prévisions est peu fournie, au contraire de celle étudiant leur qualité, nous souhaitons établir des liens entre qualité et valeur. S'ils peuvent être généralisés, ceux-ci permettraient de démontrer et de comprendre la valeur ajoutée des prévisions probabilistes.

Dans une première étape, un cadre d'évaluation de la qualité des prévisions probabilistes est présenté. Il est constitué d'un ensemble de métriques, d'outils graphiques et de modèles de référence. À l'issue d'une revue bibliographique qui recense différentes métriques probabilistes, nous proposons un nouvel outil appelé matrice d'espérance qui a pour but de comparer celles-ci. Il permet dans un premier temps de sélectionner les scores les plus adaptés pour répondre aux questions de recherche, et dans un second temps de quantifier les gains potentiels de qualité que l'utilisation de prévisions probabilistes permet théoriquement d'obtenir. Pour compléter ce cadre de référence, nous introduisons un nouveau modèle de référence que nous suggérons d'appeler CSD-CLIM. Il est construit sur l'idée d'utiliser au mieux toute l'information contenue dans les données historiques d'observations d'un site pour concevoir une prévision fiable, de qualité supérieure à celle des modèles de référence communément utilisés. Ensuite, les gains réellement obtenus que l'on peut effectivement attribuer à l'utilisation de l'approche probabiliste sont calculés pour plusieurs modèles de prévision probabiliste couramment utilisés et pour un ensemble de six sites présentant différentes conditions climatologiques. Ce calcul est effectué à l'aide d'une nouvelle méthodologie nommée "MC-CRPS", pour des prévisions J+1 construites par post-traitement statistique de prévisions d'ensemble.

Une seconde étape est dédiée à l'étude de la valeur des prévisions du point de vue de l'utilisateur. La valeur d'une prévision dépend de son utilisation, il est donc nécessaire de choisir un cas d'application pratique pour la mesurer. Pour un acteur industriel, elle mesure le bénéfice apporté par l'utilisation de cette prévision dans un contexte de prise de décision. Une recherche théorique est conduite sur la manière de modéliser les coûts associés aux décisions prises à l'aide de la prévision. On parvient à proposer une méthodologie d'évaluation de la valeur dans un cadre d'utilisation bien défini. Celle-ci recommande l'utilisation d'un graphique combinant deux informations indépendantes, à savoir la structure des coûts et les performances de la prévision pour tous les niveaux de probabilité. Cet outil permet donc de décorréler les facteurs explicatifs provenant de la qualité de la prévision et du contexte du marché.

Cette méthodologie est finalement utilisée pour expliquer les résultats obtenus lors de l'utilisation de différentes prévisions pour la valorisation de la production électrique d'une

centrale solaire sur trois marchés de l'électricité. Les résultats prouvent que la structure des prix du marché a bien un rôle fondamental sur la valeur effective des prévisions. À nouveau, les gains de l'approche probabiliste par rapport à l'approche déterministe sont évalués puis comparés aux gains de qualité. Les valeurs des prévisions probabilistes surpassent celles des prévisions déterministes. Dans un contexte d'utilisation industrielle où la structure de coûts devient elle-même incertaine, il est nécessaire de disposer également d'une prévision des prix du marché. Une méthode simple a été implémentée. Cette incertitude supplémentaire entraîne une diminution de la différence de valeur entre prévisions probabilistes et déterministes.

Extended abstract

Solar forecasting is a major asset in many solutions for large-scale integration of solar power into electricity mixes. This work aims to prove that probabilistic forecasting, yet not frequently used at an industrial scale, is nonetheless for many applications. Since the scientific literature related to value of forecasts is rare, contrary to literature about quality, this work is devoted to establishing links between quality and value. If they can be generalized, they would demonstrate and help the understanding of added-value of probabilistic forecasting.

In a first step, a framework for the assessment of quality of probabilistic forecasts is presented. It is constituted of metrics, graphic tools and reference benchmark models. After a bibliographic review focused on probabilistic metrics, a new tool called “outcome matrix” designed for the comparison of probabilistic metrics is presented. It is firstly used to select the most appropriate scores for the research question, and then to quantify the potential benefits in terms of quality that the usage of probabilistic forecasting may bring. In order to complete the framework of quality assessment of probabilistic forecasts, we propose a new reference model to benchmark solar probabilistic forecasts called “CSD-CLIM”. The main idea is to use at best all knowledge contained in the historical data of a site to generate a reliable forecast. The quality of this model slightly outperforms those of commonly used reference models. Finally, the benefits really obtained from the use of probabilistic forecasts are calculated for several probabilistic forecasting models from the literature and on six sites facing very different climatologies. This calculation is made with a new methodology called “MC-CRPS”, for day-ahead forecasts built by post-processing of ensemble forecasts provided by ECMWF.

A second step is dedicated to the study of the value of forecasts from a user perspective. The value of a forecast depends on its usage, and it is then necessary to select a practical case study to quantify it. For industrials, it measures the gains brought by this forecast in a decision-making process. A theoretical research is conducted on the way to model the cost associated to decisions based on the forecast. A methodology for the evaluation of the value of a probabilistic forecast of continuous variables is proposed. It consists in the usage of a plot that combines two independent informations, namely the cost structure and the skills of the forecast for all the different probability levels. This methodology allows to consider separately the factors explicable by the quality of the forecast and by the market cost structure.

Finally, this methodology is used for the analysis of the results obtained by different forecasting models for the economic optimization of the production of a solar powerplant on three different electricity markets. The results tend to prove that the cost structure has indeed a crucial role on the effective economic value of forecasts. Again, the gains of the probabilistic approach compared to the deterministic approach are evaluated and compared with quality gains. For an industrial usage, it must be noted that it is necessary

to have also a forecasting model dedicated to market prices. A simple method of such a forecast has been implemented. On the considered cases, the values of the probabilistic forecasts outpace those of deterministic forecasts. However, the differences between these two approaches decrease as soon as the cost structure is itself uncertain.

Remerciements

Je me dois de débiter ces remerciements par mon équipe d'encadrement, Philippe Lauret et Mathieu David, qui a su me laisser plus libre dans mes recherches que je n'aurais pu l'espérer. Je pense que c'est grâce à l'équilibre que nous avons trouvé que le travail final reflète fidèlement le parcours de recherche que j'ai souhaité suivre. Ils ont également toujours répondu présent avec bienveillance lorsque j'avais besoin de soutien, et m'ont orienté lorsque je m'engageais dans des directions de recherche peu prometteuses.

Je souhaite également exprimer ma reconnaissance auprès de M. Stéphane Grieu et M. Richard Perez. Leur travail consciencieux de relecture a permis à mon travail de thèse de s'enrichir de leurs points de vue extérieurs, et d'être sans aucun doute plus complet et plus cohérent. Je tiens à remercier ensuite les examinateurs, M. Philippe Blanc, M. Jordi Badosa, M. Pierre Pinson, qui ont tous, par leurs questions et remarques constructives, animé des discussions passionnantes à la suite de mon exposé. Je remercie également ces deux derniers pour leur accueil au sein de leurs laboratoires respectifs. Ces deux stages de recherche ont contribué grandement à la construction de mon travail. J'ai également une pensée pour l'équipe des doctorants du DTU, et pour Thibaut, vaillant adversaire échiquéen. Je n'oublie bien sûr pas que M. Rodrigo Alonso Suarez et M. Emeric Tapaches ont eu un rôle important par le temps qu'ils ont consacré à mes comités de suivi de thèse et par leurs précieux conseils.

Je salue aussi Hèlène, Maareva et Christophe qui ont été les représentants des doctorants du laboratoire PIMENT pendant mes 3 années de travail. À travers eux, je remercie l'ensemble des doctorants du laboratoire PIMENT que j'ai pu croiser pendant mon travail. J'ai une pensée spéciale pour Faly, mon fidèle compagnon de bureau. Il me faut aussi remercier les amis de Matméca, bien que je regrette que les années passées ne nous ont pas permis de nous voir très souvent. Les discussions avec certains d'entre eux ont pesé dans mon choix de m'engager dans cette thèse, et peut-être que sans eux ce projet n'aurait pas existé! Merci donc à Etienne, Charlotte, Sébastien, Anne, Chris, Laura, Flo, Marion, François, Kelly, Mattmau, Flore, Gabi, Valou, Pignet, Amiard. J'ai une pensée également pour Flo et Saspì, leurs encouragements et leurs conseils avisés m'ont été utiles plus d'une fois.

Il vient le temps de remercier chaleureusement ma famille. Mes parents m'ont toujours encouragé dans mes projets et conseillé. Les skype-jeux avec mes frères et soeur Guénolé, Yves-Maël, Adrien et Isabelle ont été des parenthèses salutaires pour mon esprit. Ensuite, je me dois d'exprimer ma gratitude auprès de ma famille réunionnaise. Tous m'ont aidé à leur manière, Antho et Julie comme traducteurs émérites, Thomas, dont les dessins de mamies m'ont permis de faire bonne figure à MT180, Linda et Karl, par leurs encouragements, et Ginger, par son énergie.

Enfin, je ne pourrais pas remercier Emma suffisamment pour tout le soutien qu'elle m'a apporté durant ces trois longues années de thèse. J'ai conscience des sacrifices qu'elle

aussi a dû consentir pour m'aider à mener à bien ce travail qu'elle n'a pourtant pas choisi. Elle m'a aidé à surmonté les difficultés, m'a épaulé pour la rédaction, la relecture des publications et du manuscrit. Je dois également mentionner son talent en design graphique; derrière de nombreux graphiques ou schémas de ce manuscrit se cache ses idées et son imagination.

Table des matières

0	Introduction	1
0.1	Contexte général	2
0.2	Les caractéristiques de l'énergie solaire	4
0.3	Les problématiques liées à l'utilisation à grande échelle de l'énergie solaire	6
0.4	Solutions proposées par la communauté scientifique	7
0.5	Différents types de prévisions solaires pour différents usages	7
0.6	Contexte scientifique spécifique de la thèse : objectifs et démarche	11
0.7	Structure du manuscrit	13
1	Évaluation de la qualité des prévisions probabilistes	15
1.1	Introduction	17
1.2	Notions préliminaires	19
1.3	Les attributs relatifs à la qualité d'une prévision probabiliste	20
1.4	Outils d'évaluation de la qualité	22
1.5	Comparaison des qualités des prévisions déterministes et probabilistes	37
1.6	Conclusions	51
	Annexes	53
1.A	Démonstration du caractère propre des scores binaires de type "fonction puissance"	53
1.B	Décomposition du <i>Brier score</i>	54
1.C	Équivalences du <i>Brier score</i> , du <i>score propre linéaire</i> et du <i>Winkler score</i>	56
2	Un nouveau modèle de référence climatologique pour les prévisions solaires probabilistes	57
2.1	Introduction	59
2.2	Required properties for a good climatology benchmark model	61
2.3	Context of the study	62
2.4	Climatology benchmark models for solar probabilistic forecasting	66
2.5	Results	70
2.6	Discussion	75
2.7	Conclusions	79
	Appendices	81
2.A	Sensitivity Analysis related to the number of bins used by the CSD-CLIM binning process	81
2.B	Time-invariance of CSD-CLIM	82
2.C	Reliability assessment using PIT Histograms	83
3	Gains effectifs des approches probabiliste et ensembliste	85

3.1	Introduction	87
3.2	Building probabilistic forecasts	90
3.3	Verification of the probabilistic forecasts	96
3.4	Case studies	100
3.5	Results	101
3.6	Discussion	106
3.7	Conclusions	112
	Appendices	113
3.A	Data quality check	113
3.B	Bias and standard deviation of EPS members distribution and observations for the six sites	116
3.C	Selection of the optimal α	118
3.8	Compléments	120
4	Valeur des prévisions probabilistes et liens avec la qualité	123
4.1	Introduction	125
4.2	Valeur des prévisions de variables binaires	126
4.3	Une méthodologie pour l'évaluation de la valeur des prévisions de variables continues	128
4.4	Application à un cas pratique : Valorisation économique d'un parc de pro- duction solaire	141
4.5	Conclusions	172
	Annexes	174
4.A	La méthode "Variance Deficit"	174
4.B	Modélisation de la production photovoltaïque	175
4.C	Graphique de <i>quantile score</i> d'une prévision déterministe	180
4.D	Mise en place du modèle de prévision des prix ARIMA	180
	Conclusions générales et perspectives	184
	Résultats	184
	Perspectives de travail	184

Liste des figures

1	Consommation énergétique mondiale par source d'énergie primaire.	2
2	Production électrique de l'île de la Réunion en 2019.	4
3	Décomposition du rayonnement solaire.	5
4	Classification des différentes familles de prévisions solaires.	8
5	Fonctionnement schématique d'une prévision d'ensemble.	9
6	Objectif et démarche du travail.	13
7	Organisation schématique du manuscrit.	14
1.1	Prévision déterministe et prévision probabiliste.	18
1.2	Diagrammes de rang.	34
1.3	Diagramme de fiabilité.	35
1.4	Diagramme PINAW.	36
1.5	Diagramme de calibration.	36
1.6	Matrice d'espérance du <i>score linéaire naïf</i>	40
1.7	Décomposition du <i>Brier score</i> sur la matrice d'espérance.	40
1.8	Matrice d'espérance du <i>Brier score</i> pour une prévision réelle.	41
1.9	Matrices d'espérance des scores de prévisions binaires.	42
1.10	Gains potentiels des scores de prévisions binaires.	43
1.11	Matrices d'espérance des scores de prévisions de variables continues.	45
1.12	Gains potentiels des scores de prévisions de variables continues.	46
1.13	Fonctions CDF des prévisions du cas d'étude.	48
1.14	Amélioration de la fiabilité.	49
1.15	Amélioration de la résolution.	50
1.16	<i>CRPS</i> , incertitude, fiabilité, résolution et finesse du cas d'étude.	51
2.4.1	Uncertainty of the <i>CRPS</i>	66
2.4.2	Construction of <i>CSD-UNC</i>	69
2.4.3	Comparison between <i>UNC</i> and <i>CSD-UNC</i>	69
2.4.4	Histograms of k_t^* for Marshall Island and Desert Rock.	70
2.5.1	<i>CRPS</i> of the benchmark models vs. simple post-processing.	74
2.6.1	Decomposition of <i>CSD-UNC</i> according to clearsky irradiance bins.	75
2.6.2	$UNC^{i=24}$ for a specific bin.	76
2.6.3	Contingency tables related to the two binning processes for three sites.	77
2.6.4	Differences in the binning process between CH-PeEn and CSD-CLIM.	78
2.6.5	Clear-sky index distributions for Toravere.	79
2.A.1	Impact of Nb on <i>CSD-UNC</i>	81
2.C.1	PIT Histograms of the three benchmark models.	84
3.2.1	Uniform construction of a CDF.	91
3.2.2	Observed GHI and the predicted day-ahead GHI.	92

3.2.3	Distributions of observations for a fixed level of forecasting.	94
3.3.1	Virtual forecasts related to MC-CRPS.	99
3.5.1	<i>CRPS skill score</i> of each model for the six considered sites.	102
3.5.2	MC-CRPS of the six sites and all forecasting models.	105
3.6.1	Standard deviation of EPS members and observations.	107
3.6.2	Accuracy of the second moment.	109
3.6.3	Link between G_2 and the accuracy of the second moment.	111
3.A.1	Quality check for Desert Rock.	113
3.A.2	Quality check for Saint-Pierre.	114
3.A.3	Quality check for Hawaii.	114
3.A.4	Quality check for Palaiseau.	115
3.A.5	Quality check for Tiruvallur.	115
3.A.6	Quality check for Langley.	116
3.B.1	Bias and standard deviation of EPS members and observations 1/2.	117
3.B.2	Bias and standard deviation of EPS members and observations 2/2.	118
3.C.1	Determination of α for the analog ensemble technique.	119
3.8.1	<i>CRPS skill scores</i> par rapport au modèle CSD-CLIM ₃₀	120
3.8.2	Gains des approches probabilistes et ensemblistes.	122
4.2.1	Exemple d'un graphique de PEV.	128
4.3.1	Trois exemples de fonctions pertes.	130
4.3.2	Minimisation des pertes pour une prévision déterministe.	131
4.3.3	minimisation des pertes pour une prévision d'ensemble.	132
4.3.4	<i>Quantiles scores</i> pour différents modèles de prévision.	135
4.3.5	PEVC de 3 modèles de prévision.	136
4.3.6	Graphique de répartition des risques.	137
4.3.7	Graphiques EVC du modèle VD.	138
4.3.8	Distribution uniforme.	139
4.3.9	Distribution centrée.	140
4.4.1	Prix du marché et volume échangé dans un marché à fixing.	144
4.4.2	Processus temporel du trading SPOT d'EPEX.	144
4.4.3	"Merit order" d'un marché à fixing.	145
4.4.4	Contraintes sur les prévisions utilisables sur les marchés J+1.	146
4.4.5	Modélisation des pénalités pour un producteur d'électricité.	147
4.4.6	Gains pour un producteur en fonction de sa déviation D	148
4.4.7	Schéma du processus d'optimisation d'un moyen de production solaire.	149
4.4.8	Fonction perte pour l'objectif O_1	150
4.4.9	Fonction perte pour l'objectif O_2	151
4.4.10	Comparaison des fonctions \hat{P}_1 et \hat{P}_2 des objectifs O_1 et O_2	152
4.4.11	Prévisions du ratio R	154
4.4.12	Distribution du ratio R pour les 3 pays.	156
4.4.13	Résultats pour l'objectif O_1 sur le marché français.	158
4.4.14	Résultats pour l'objectif O_1 sur le marché portugais.	159
4.4.15	Résultats pour l'objectif O_1 sur le marché suisse.	160
4.4.16	Résultats pour l'objectif O_2 sur le marché français.	162
4.4.17	Résultats pour l'objectif O_2 sur le marché portugais.	163
4.4.18	Résultats pour l'objectif O_2 sur le marché suisse.	164

4.4.19	Graphiques EVC des prévisions “control” et VD	166
4.4.20	Graphiques EVC des prévisions ECMWF brutes et <i>NR GEV</i>	167
4.4.21	Pénalité moyenne en fonction de l’erreur de prévision du ratio R	169
4.4.22	Corrélation entre <i>quantile score</i> et coût moyen.	170
4.4.23	Gains maximums de valeur et de qualité de l’approche probabiliste.	171
4.B.1	Processus de modélisation de la production.	176
4.B.2	Suréchantillonnage linéaire classique de l’éclairage énergétique.	178
4.B.3	Suréchantillonnage linéaire avec correction de consistance.	179
4.D.1	Coefficients d’autocorrélation de (R, γ) pour la France.	181
4.D.2	Coefficients d’autocorrélation de (R, γ) pour le Portugal.	181
4.D.3	Coefficients d’autocorrélation de (R, γ) pour la Suisse.	181

Liste des tableaux

1	Différents usages des prévisions solaires.	10
1.1	Carte des scores propres probabilistes.	31
1.2	Outils graphiques de diagnostic des prévisions probabilistes.	37
1.3	Interprétation de la matrice d'espérance d'un score.	39
1.4	Caractéristiques des mesures de GHI de Saint-Pierre, années 2012 et 2013.	47
1.5	Présentation des caractéristiques des 8 prévisions du cas d'étude.	48
1.6	Synthèse des gains potentiels maximums de l'approche probabiliste	52
2.2.1	Properties of a good climatology benchmark model.	62
2.3.1	Characteristics of sites used for the study.	63
2.4.1	Example of the binning process of CSD-CLIM.	68
2.5.1	Decomposition of the <i>CRPS</i> of the 3 tested benchmark models.	72
2.5.2	<i>CRPS</i> (W/m^2) of the LQR forecast (grey) and the 3 benchmark models.	73
2.5.3	Overall qualitative assessment of the benchmark models.	75
2.B.1	Periods used for time stability assessment.	82
2.B.2	Data lengths used for time stability assessment.	82
2.B.3	Sensitivity of <i>CSD-UNC</i> on data period.	82
2.B.4	Sensitivity of <i>CSD-UNC</i> on data length.	83
3.2.1	<i>RMSE</i> of 4 deterministic forecasts that can be inferred from an EPS.	95
3.2.2	Summary of all considered forecasting models.	96
3.4.1	Main characteristics of time series of GHI used to test the models.	101
3.5.1	<i>CRPS</i> and its components for all considered models.	104
3.8.1	Gains probabilistes effectifs de qualité.	122
4.2.1	Table de contingence d'une prévision binaire.	127
4.3.1	Les modèles de prévision utilisés	129
4.3.2	Outils d'étude de la valeur d'une prévision probabiliste.	141
4.4.1	<i>RMSE</i> des prévisions des variables de prix.	155
4.4.2	Variance des variables de prix.	155
4.4.3	Tableau récapitulatif des performances des différentes prévisions.	168
4.4.4	Gains probabilistes effectifs de valeur.	170
4.4.5	Gains de qualité et de valeur de l'approche ensembliste.	172
4.B.1	Spécifications techniques des panneaux photovoltaïques.	177
4.D.1	Dépendance du ratio R avec les variables temporelles.	182
4.D.2	Variables exogènes retenues pour la désaisonnalisation des variables.	182

Notations

X	Variable aléatoire cible (ici le GHI)
\hat{F}_d	Prévision déterministe
ε	Erreur d'une prévision déterministe
O	Observation
p	Probabilité
(\hat{F}_0, \hat{F}_1)	Prévision binaire
(Q_0, Q_1)	Probabilités réelles d'une variable binaire
\overline{O}	Climatologie d'une variable binaire
$(\hat{F}_{i,i=1\dots I})$	Prévision multi-catégories
$(Q_{i,i=1\dots I})$	Probabilités réelles d'une variable multi-catégories
\hat{f}	Fonction de densité associée à une prévision probabiliste
q	Fonction de densité réelle d'une variable aléatoire
\hat{F}	Fonction de répartition associée à une prévision probabiliste
\mathcal{Q}	Fonction de répartition réelle d'une variable aléatoire
\mathcal{O}	Fonction de répartition associée à une observation
$\overline{\mathcal{Q}}$	Fonction de répartition associée à la climatologie (variable continue)
$(E_{i,i=1\dots M})$	Prévision d'ensemble
τ	Niveau de probabilité
\hat{F}_τ	Prévision du quantile associé à un niveau de probabilité τ
Q_τ	Quantile associé à un niveau de probabilité τ
N	Nombre de couples prévision/observation d'un système de prévision
N_A	Nombre d'observations d'une période d'apprentissage
\mathbb{E}	Espérance
H	Fonction de Heaviside
$\mathbf{1}$	Fonction indicatrice
μ	Moyenne d'une distribution
σ	Écart-type d'une distribution

Chapitre 0

Introduction

Sommaire

0.1	Contexte général	2
0.2	Les caractéristiques de l'énergie solaire	4
0.3	Les problématiques liées à l'utilisation à grande échelle de l'énergie solaire	6
0.4	Solutions proposées par la communauté scientifique	7
0.5	Différents types de prévisions solaires pour différents usages	7
0.6	Contexte scientifique spécifique de la thèse : objectifs et démarche . .	11
0.7	Structure du manuscrit	13

0.1 Contexte général

La consommation d'énergie mondiale est en constante augmentation depuis les années 1970. Celle-ci a presque triplé en 50 ans (*Key World Energy Statistics*, AIE, 2017), pour atteindre environ 600 exajoules en 2019. Même si certaines prévisions anticipent un léger recul de la demande en 2021 à cause de l'épidémie mondiale de COVID-19, cette tendance de long terme devrait perdurer (*World Energy Outlook*, AIE, 2020). La figure 1 montre l'évolution de cette consommation depuis 1994 et les différentes sources d'énergie primaire utilisées pour répondre à la demande.

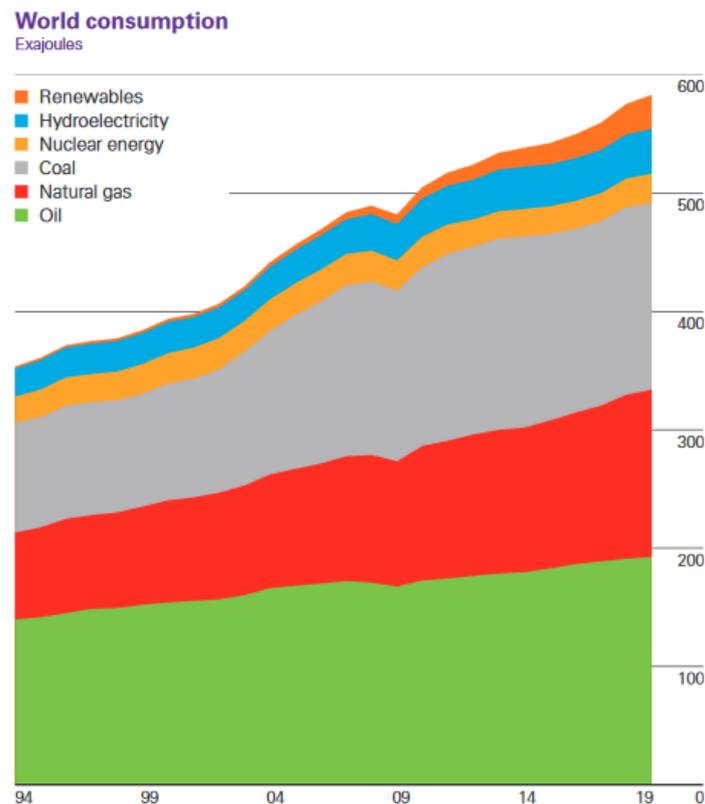


FIGURE 1 : Consommation énergétique mondiale par source d'énergie primaire de 1994 à 2019, source : *Statistical Review of World Energy*, 2020.

Le pétrole et le charbon restent les énergies primaires principales. Ce sont aussi celles qui absorbent aujourd'hui majoritairement la croissance de la consommation énergétique mondiale. Cependant, au rythme actuel d'utilisation, la société BP estime l'épuisement des stocks de charbon dans 130 ans et des stocks de pétrole et de gaz naturel dans 50 ans environ (*Statistical Review of World Energy*, BP, 2020). Dans le but d'anticiper cet épuisement, le modèle énergétique mondial actuel est donc appelé à se transformer dans les années à venir.

En parallèle, durant ces dernières décennies, l'accélération du réchauffement climatique s'est imposée comme un sujet de préoccupation politique majeur. En 1992, le sommet de la Terre à Rio de Janeiro s'achève par la signature de la convention-cadre des Nations Unies sur les changements climatiques, qui devient alors le premier accord international traitant de ce sujet. Plusieurs autres sommets internationaux suivront, avec pour objectif principal de réduire les causes anthropiques du réchauffement climatique.

La COP (“Conference Of the Parties”) de décembre 2015, et l’“Accord de Paris” qui en a été la conséquence, ont scellé l’ambition des parties de contenir ce réchauffement en limitant l’augmentation de la température mondiale sous le seuil de 2°C. Or, le réchauffement climatique d’origine anthropique est principalement causé par l’émission de gaz à effet de serre, et le secteur énergétique est particulièrement concerné. Au sein de ce secteur, la production d’électricité occupe une place non négligeable (Abdallah et El-Shennawy, 2013). Le pétrole est relativement peu utilisé pour la production d’électricité, mais Schlömer et al. (2014) montrent que les centrales thermiques utilisant le charbon, le fuel ou le gaz naturel sont les plus émettrices de gaz à effet de serre.

Pour diminuer l’utilisation de ces énergies fossiles polluantes pour la production d’électricité, il convient de changer les usages et de recourir à l’utilisation d’énergies de substitution. Parmi celles-ci, certaines sont encore peu utilisées aujourd’hui mais possèdent un fort potentiel de déploiement. On peut citer l’énergie nucléaire ou les énergies renouvelables (hydraulique, éolienne, solaire, marines, géothermique, etc.). Si elle est n’émet pas de gaz à effet de serre, l’utilisation massive de l’énergie nucléaire entraîne de fortes problématiques de gestion et de stockage des déchets. S’ajoute à cela une considération stratégique. Il peut être problématique pour un territoire de devenir dépendant de l’approvisionnement en uranium, nécessaire à l’exploitation de l’énergie nucléaire ; et ce d’autant plus pour le secteur énergétique, particulièrement sensible. En outre, les stocks d’uranium sont eux aussi limités. C’est pourquoi l’accord de Paris s’est décliné en une série d’objectifs, dont l’un des principaux est de porter à 32% la part des énergies dites “renouvelables” dans la consommation énergétique finale en 2030 (Commission Européenne, 2018).

Toutes ces raisons sont à l’origine d’une volonté d’intégration massive des énergies renouvelables au sein des réseaux électriques. Comme l’utilisation de l’énergie hydraulique est limitée par le potentiel hydrique des territoires, que l’exploitation de l’énergie géothermique n’est efficace à large échelle que dans certains territoires notamment volcaniques, et que les énergies marines sont globalement encore à un stade précoce de développement, les énergies éoliennes et solaires font l’objet d’un intérêt croissant. Cette tendance est assez globale et les études s’accordent à dire qu’elle pourrait s’accélérer, certaines suggérant même que la capacité installée de production solaire et éolienne pourrait doubler d’ici 2030 (Li, 2017). Il faut cependant noter que le potentiel de ces deux sources d’énergie est très dépendant de la situation géographique. À titre d’exemple, le Danemark a su tirer profit des forts vents de la mer du Nord pour développer considérablement son parc éolien. Celui-ci assure aujourd’hui plus de 50% de la production électrique du Danemark. Ainsi, chaque territoire sélectionne la ressource la mieux adaptée à son contexte géographique.

Diagne (2015) a étudié le cas particulier de l’île de la Réunion. La figure 2 le résume, en présentant le bilan de sa production électrique en 2019.

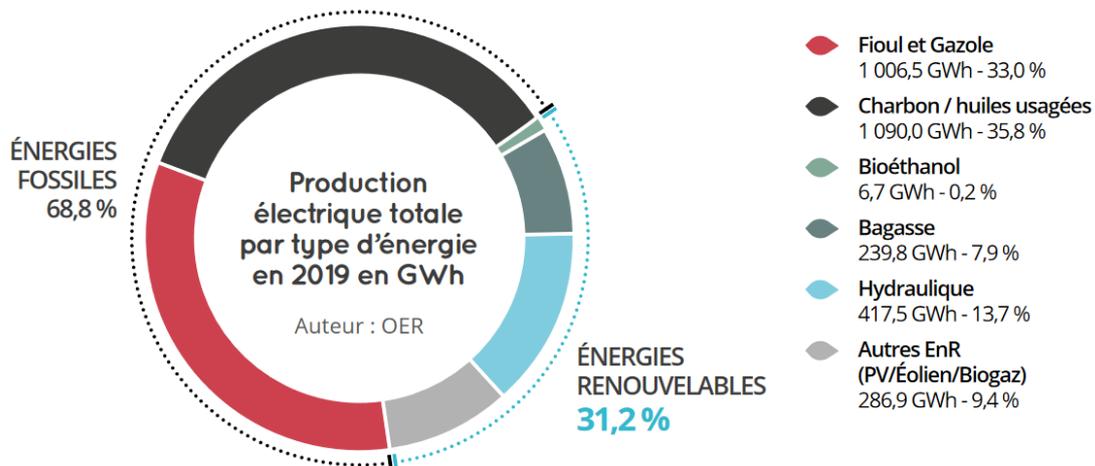


FIGURE 2 : Production électrique de l'île de la Réunion en 2019, source : *Bilan Énergétique de la Réunion 2019*, Observatoire Énergie Réunion, 2020.

Il apparaît clairement que les énergies fossiles sont encore majoritaires. C'est problématique au regard des objectifs de réduction d'émission de gaz à effet de serre. De plus, ces ressources sont presque exclusivement importées, ce qui engendre une dépendance énergétique du territoire. Pour répondre à ces enjeux, le plan de programmation pluriannuel de l'énergie ("PPE") prévoit un développement ambitieux des énergies renouvelables. Parmi ces énergies, le bioéthanol et la bagasse sont encore minoritaires et l'énergie hydraulique n'a plus un potentiel de déploiement important. Située proche du tropique du Capricorne, l'île de la Réunion dispose d'un ensoleillement important ; elle a donc fait le choix de privilégier l'énergie solaire. Pour preuve, sa puissance installée de production électrique photovoltaïque est aujourd'hui dix fois supérieure à celle de l'énergie éolienne (*Bilan Énergétique de la Réunion 2019*, Observatoire Énergie Réunion, 2020). La PPE prévoit d'ailleurs une accélération du développement de l'énergie solaire dans les années à venir (*Programmation pluriannuelle de l'énergie de La Réunion*, Journal Officiel, 2017). Se pose alors la question de la possibilité technique de l'augmentation de la part de l'énergie solaire dans un réseau électrique comme celui de la Réunion.

0.2 Les caractéristiques de l'énergie solaire

Dans la suite, les termes "rayonnement solaire" et "éclairage énergétique" désignent la puissance surfacique du rayonnement solaire et sont considérés équivalents. Les unités de production électrique utilisant l'énergie solaire ont recours majoritairement à des cellules photovoltaïques. Ce sont des composants électroniques qui ont la propriété de produire de l'électricité lorsqu'ils sont exposés à un éclairage énergétique. La puissance électrique résultante est presque proportionnelle à la puissance de l'éclairage énergétique reçu sur la surface. Elle dépend aussi du rendement de la cellule, un paramètre variable, notamment en fonction de la température de la cellule. La production d'une cellule orientée horizontalement et soumise au rayonnement solaire dépend alors directement de l'éclairage énergétique horizontal (ou "GHI") (Lave et al., 2013). Un module photovoltaïque est simplement l'agrégation de nombreuses cellules reliées entre elles en série et en parallèle. Les panneaux solaires photovoltaïques peuvent être utilisés sur de nombreux supports :

de l'échelle individuelle avec le panneau sur toiture aux grandes centrales solaires connectées au réseau de transport d'électricité. L'inclinaison des panneaux solaires est variable et déterminée soit en fonction des contraintes intrinsèques du site d'implantation (cas des toitures avec une pente préexistante) soit dans un objectif de maximisation de la production annuelle. Dans ce cas, la production ne dépend plus directement de l'éclairement horizontal. Néanmoins, et sachant que l'éclairement énergétique sur un plan incliné ("GTI") est étroitement lié au GHI, seul l'éclairement horizontal est pris en compte dans ce travail. De plus, Vallance (2018) montre que l'étude du GHI assure une plus grande universalité des résultats et des conclusions.

Le GHI est ainsi la puissance totale incidente de rayonnement sur une surface horizontale à un instant donné, et c'est le principal facteur régissant la production d'un panneau photovoltaïque. Il est relativement simple de calculer l'énergie de rayonnement incident au sommet de l'atmosphère. Cependant, au cours de la traversée de l'atmosphère, ce rayonnement va s'affaiblir. Une partie du rayonnement va être piégée (diffusée ou réfléchie durant la traversée, ou à l'interface des différents milieux composant l'atmosphère). Cette part piégée du rayonnement peut s'échapper, ou rester présente dans l'atmosphère. Dans ce cas, elle constitue ce qu'on appelle le rayonnement diffus. Ce processus de décomposition du rayonnement est illustré sur la figure 3.

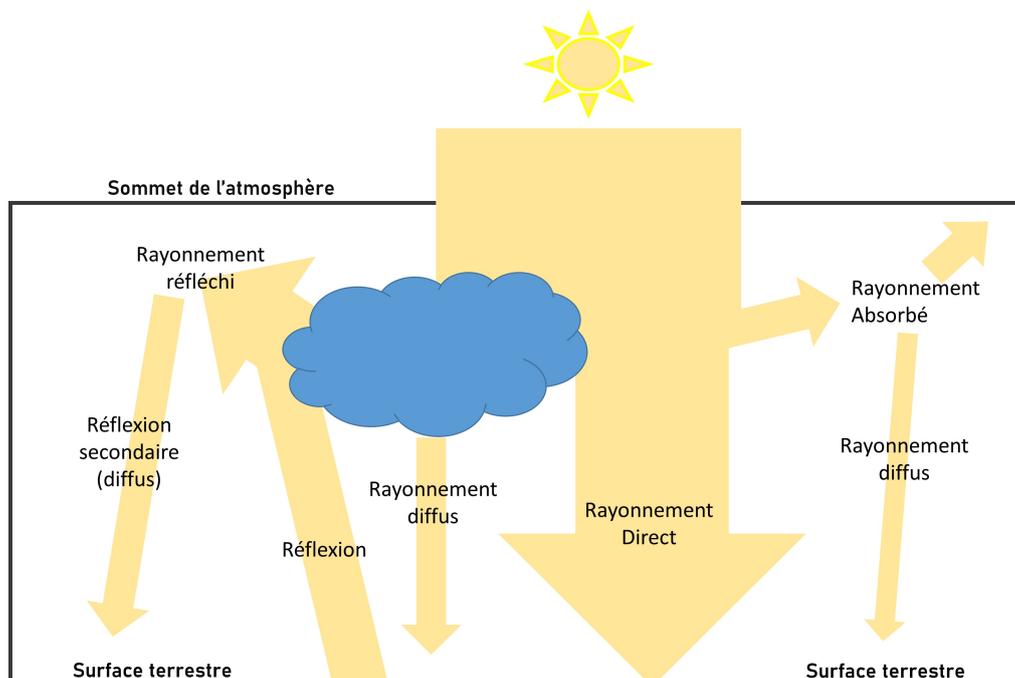


FIGURE 3 : Illustration de la décomposition du rayonnement au cours de la traversée de l'atmosphère.

De cette façon, le GHI incident sur un panneau est la somme de deux termes : le rayonnement solaire direct (BHI pour "Beam Horizontal Irradiance") et le rayonnement solaire diffus (ou DHI : "Diffuse horizontal irradiance"), auxquels on peut ajouter éventuellement le rayonnement réfléchi par l'environnement proche via l'albédo. En l'absence de couverture nuageuse, le GHI au sol dépend principalement de l'angle solaire et de la quantité de vapeur d'eau et d'aérosols présents dans l'atmosphère. Des modèles nommés modèles de ciel clair permettent de calculer ce rayonnement avec une marge d'erreur relativement

faible. Les rayonnements (dits “de ciel clair” ainsi déterminés et notés $\text{GHI}_{\text{ciel clair}}$) ont de fortes variabilités temporelles, à différentes échelles. Il existe par exemple une variabilité journalière correspondant à l’alternance jour/nuit, ou encore des variabilités saisonnières dues à l’inclinaison de l’axe de rotation de la Terre. À cette composante déterministe de ciel clair vient s’ajouter une composante beaucoup moins prédictible due à la couverture nuageuse qui peut agir comme un masque et réduire drastiquement la puissance du rayonnement solaire. L’atténuation du rayonnement de ciel clair par les nuages peut alors être mesurée en calculant l’ “indice de ciel clair” k_t^* défini par :

$$k_t^* = \frac{\text{GHI}}{\text{GHI}_{\text{ciel clair}}}. \quad (1)$$

0.3 Les problématiques liées à l’utilisation à grande échelle de l’énergie solaire

On qualifie ainsi l’énergie solaire, au même titre que l’énergie éolienne, d’énergie “intermittente fatale”, c’est-à-dire que sa disponibilité est non contrôlable et imposée par la valeur instantanée de la ressource. En outre, cette ressource affiche une forte variabilité spatio-temporelle. De plus, elle n’est pas, en général, en phase avec les besoins en énergie des consommateurs : elle peut être absente lors d’appels d’énergie sur le réseau ou au contraire présente alors que les besoins sont réduits. L’intégration massive de ce type d’énergie au sein du mix électrique transforme donc les défis auxquels sont confrontés les différents acteurs du secteur énergétique. Par exemple, les réseaux électriques actuels n’ont pas nécessairement été conçus pour supporter une forte pénétration d’énergie intermittente. Cette évolution de la gestion des réseaux pose plusieurs défis techniques. On peut noter par exemple les contraintes liées à la tenue en tension et à la fourniture de services systèmes des centrales photovoltaïques, la nécessité de fournir des réserves de puissance plus importantes, la difficulté à fournir des énergies de réserve et des services systèmes, ou le manque d’inertie des systèmes de production utilisant des sources d’énergies intermittentes. Ahmed et al. (2020) proposent une revue détaillée de l’ensemble de ces défis. Parmi ceux-ci, la garantie de l’équilibre temps réel entre production et consommation, couramment appelé équilibre offre/demande, nécessaire à la stabilité électrique de tout réseau, occupe une place prépondérante. En effet, le caractère variable et non contrôlable de la production renouvelable peut compliquer pour le gestionnaire de réseau les opérations de “dispatching”. Ces difficultés techniques sont d’autant plus contraignantes que le réseau est petit, et non interconnecté. En effet, un tel réseau ne dispose pas de l’inertie des grands réseaux continentaux et de leurs interconnexions, et doit être localement autonome en ce qui concerne sa production électrique. Cette fragilité du réseau se traduit en terme de gestion par des mesures visant à assurer la sécurité et la stabilité du réseau, qui peuvent ainsi venir réduire la production instantanée d’électricité solaire. Ainsi, dans les “ZNI” (Zones non interconnectées au réseau métropolitain continental, incluant les réseaux des départements et territoires d’outre-mer), dès que la part de puissance instantanée électrique issue de ressources intermittentes dépasse les 30%, les gestionnaires de ces réseaux peuvent être amenés à déconnecter des producteurs d’énergie intermittente (Arrêté du 23 avril 2008). Depuis 2017, à la Réunion, cette part relative de production électrique provenant de sources intermittentes est passée de 30% à 32%. On constate que la volonté d’augmenter la part de l’énergie solaire dans le mix électrique ne se heurte pas

au manque de gisement solaire, mais aux contraintes techniques d'intégration de cette énergie dans le réseau. Ces différentes contraintes opérationnelles sont à même de mettre à mal les objectifs de développement de l'énergie solaire détaillés dans la PPE.

0.4 Solutions proposées par la communauté scientifique

Pour aider les gestionnaires de réseau et les différents acteurs de l'énergie à relever les défis techniques liés à l'insertion massive de l'énergie photovoltaïque, la communauté scientifique a proposé différentes solutions pouvant être combinées. Certaines recherches (Hall et Bain, 2008 ; de Sisternes et al., 2016) suggèrent qu'une manière efficace de réduire les conséquences liées à l'incertitude de la production solaire est le développement coordonné du stockage et de la prévision. D'autres études ajoutent aussi le foisonnement des sources de production (Tapachès et al., 2019 ; Bridier et al., 2014 ; Schneider et al., 2008).

Perez et al. (2020) proposent quant à eux une solution combinant la prévision, le stockage, la surcapacité de production et l'écêtement pour garantir au gestionnaire de réseau un profil de production et favoriser ainsi la pénétration massive de l'énergie solaire au sein des réseaux.

Toutes ces solutions font apparaître la prévision solaire comme un élément central et crucial, et ce d'autant plus qu'elle se révèle peu coûteuse. C'est pourquoi l'obtention d'une prévision de la "meilleure qualité" (le sens donné à ce mot sera détaillé dans ce travail) possible est l'objet de nombreuses recherches scientifiques (Lorenz et al., 2009a ; Zamo et al., 2014a) dont certaines ont été réalisées en contexte insulaire (Lauret et al., 2015 ; Lauret et al., 2016). Le rôle de la prévision est d'aider à anticiper au mieux les fluctuations de la production d'énergie solaire. Dans la plupart des environnements opérationnels, les gestionnaires de réseau ont déjà intégré cette contrainte et l'ont traduite en règles, de sorte que de nombreux acteurs sont désormais concernés par la nécessité de disposer d'une prévision de production. Par exemple, en France la loi n° 2015-992 (2015) permet aux producteurs d'énergie solaire de vendre leur production sur les marchés de l'électricité, ce qui implique de disposer d'une prévision de production. En pratique, il peut être nécessaire de disposer de plusieurs modèles de prévision, pour plusieurs horizons temporels différents (journaliers, infra-journalier, infra-horaire).

0.5 Différents types de prévisions solaires pour différents usages

L'objectif de la prévision solaire est donc de fournir des outils d'aide à la décision, pour tous les acteurs impliqués dans un processus de gestion d'un réseau électrique. Ils peuvent être des gestionnaires de réseau, des producteurs, des courtiers, des gestionnaires de micro-réseau, etc. Cependant, ces acteurs peuvent avoir différents besoins en fonction de leurs objectifs, correspondant à différents horizons temporels. Il peut être nécessaire de connaître des prévisions de production d'énergie solaire à l'échelle infra-horaire (échelle de temps du recours à la réserve tertiaire rapide), à l'échelle infra-journalière (échelle de temps des aléas de production sur des groupes thermiques), ou encore journalière (planification de la production du lendemain). Or, les techniques de prévision, ainsi que les ressources né-

cessaires à leur déploiement, différent selon l’horizon temporel et la résolution temporelle souhaités. Diagne (2015) regroupe les différentes familles de techniques de prévision selon les résolutions temporelles et spatiales.

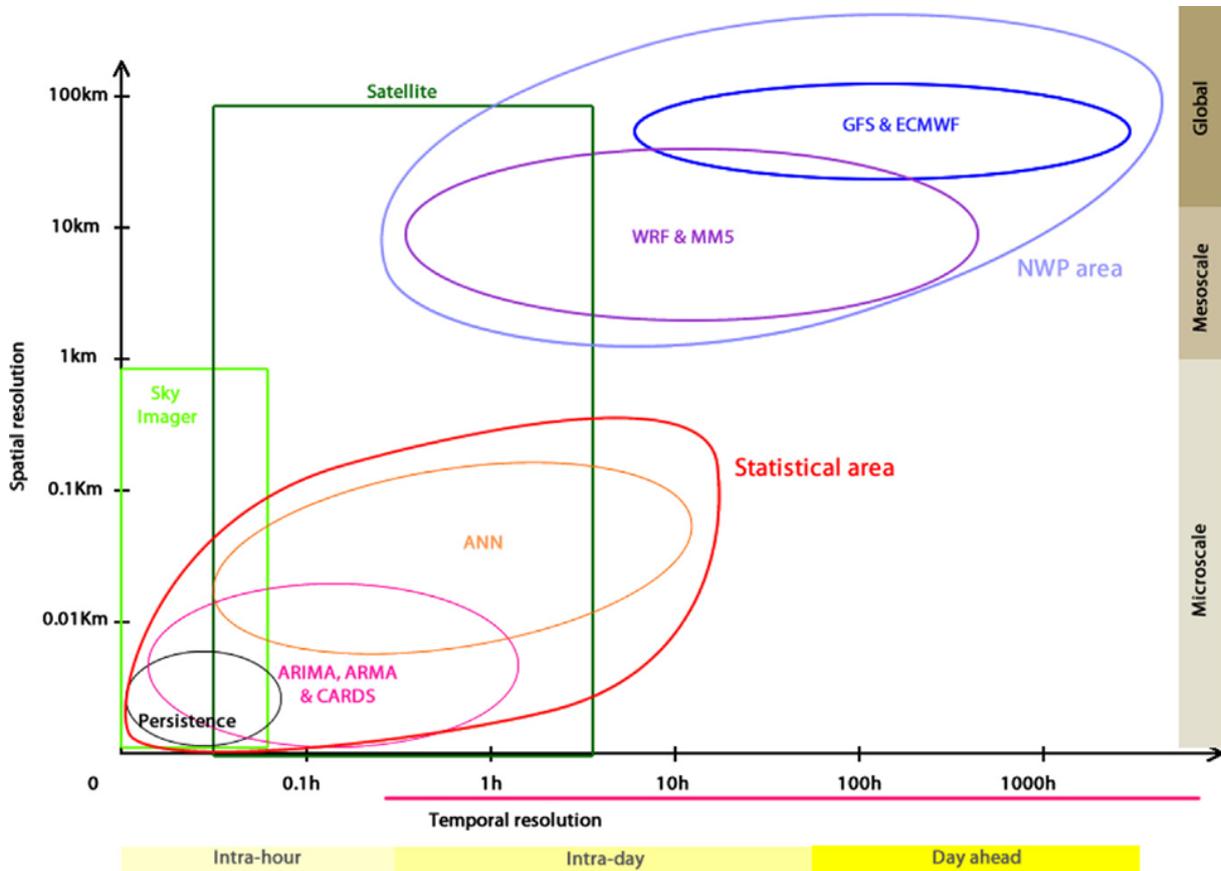


FIGURE 4 : Une classification des différentes familles de prévisions solaires, selon les résolutions spatiales et temporelles, source : Diagne (2015).

On peut distinguer trois grandes familles de techniques de prévision. Les prévisions très court-terme, (“infra-horaire” ou “intra-hour”) utilisent des méthodes statistiques s’appuyant sur des données endogènes comme les mesures passées du GHI fournies par un pyranomètre ou un sky-imager (caméra *in-situ* panoramique filmant le ciel et permettant d’anticiper le passage de la couverture nuageuse). Elles permettent d’avoir une prévision avec une résolution temporelle de l’ordre de la minute pour des horizons temporels allant jusqu’à 2 heures.

Les prévisions infra-journalières utilisent généralement le même type de techniques statistiques, mais peuvent les combiner avec des données exogènes, provenant par exemple d’imagerie satellite associées à des modèles de type “cloud-motion vector”(Escrig et al., 2013). Elles sont plutôt utilisées pour des horizons temporels inférieurs à 8 heures.

Enfin, les prévisions dites court-terme utilisent les prévisions numériques du temps (“Numerical Weather Prediction” ou NWP). Elles sont généralement utilisées pour générer des prévisions pour le lendemain (“J+1” ou “day-ahead”) et peuvent être exploitables pour un horizon temporel allant jusqu’à quelques jours. Les résolutions temporelles sont plus faibles, typiquement de l’ordre de l’heure. Ces prévisions sont issues de simulations de l’évolution de l’état de l’atmosphère, et sont effectuées par des centres de calcul mé-

téorologiques parmi lesquels on peut citer ECMWF (Europe), UKMO (Royaume-Uni), le BMRC (Australie) ou le CMA (Chine). Tous ces centres météorologiques suivent une méthodologie commune pour les calculs de simulation, appelées prévisions globales dynamiques lorsqu’elles couvrent toute la surface terrestre. Il existe aussi des modèles qui ne couvrent qu’une région de la surface terrestre (AROME de Météo France, NAM du “National Oceanic and Atmospheric Administration” aux Etats-Unis).

Des systèmes de mesure stratégiquement placés sur l’océan (bouées et navires), sur la surface terrestre (stations de mesure), dans les couches basses de l’atmosphère (radiosondes et ballons), dans les couches hautes (AIREP, données recueillies par les avions en vol) ou au-dessus de l’atmosphère (satellites) fournissent des observations caractérisant l’état de l’atmosphère. Ces observations sont assimilées par le système de calcul plusieurs fois par jour. À la fin de la période d’assimilation, ces mesures, appelées “analyse”, sont utilisées comme conditions initiales et la simulation est lancée par le serveur de calcul. Les équations utilisées par le calculateur sont résumées par Holton (2012). Parmi les centres de calcul météorologiques délivrant ce type de prévision, nous utilisons dans ce travail le centre européen ECMWF. Ce dernier propose un large éventail de prévisions. On peut classer celles-ci en 2 familles : les prévisions déterministes et les prévisions d’ensemble. Une prévision d’ensemble est un ensemble de plusieurs prévisions déterministes (51 dans le cas de la prévision d’ensemble d’ECMWF utilisée dans ce travail). Chacune représente un possible état futur de l’atmosphère. Cette prévision d’ensemble est constituée ainsi : les conditions initiales et certains paramètres du modèle sont modifiés 50 fois par rapport à l’analyse, puis la simulation est effectuée une fois pour chaque condition initiale ainsi constituée (Leutbecher et Palmer, 2008). Le membre issu de l’état initial d’origine est appelé membre de contrôle. La figure 5 présente schématiquement le fonctionnement d’une prévision d’ensemble.

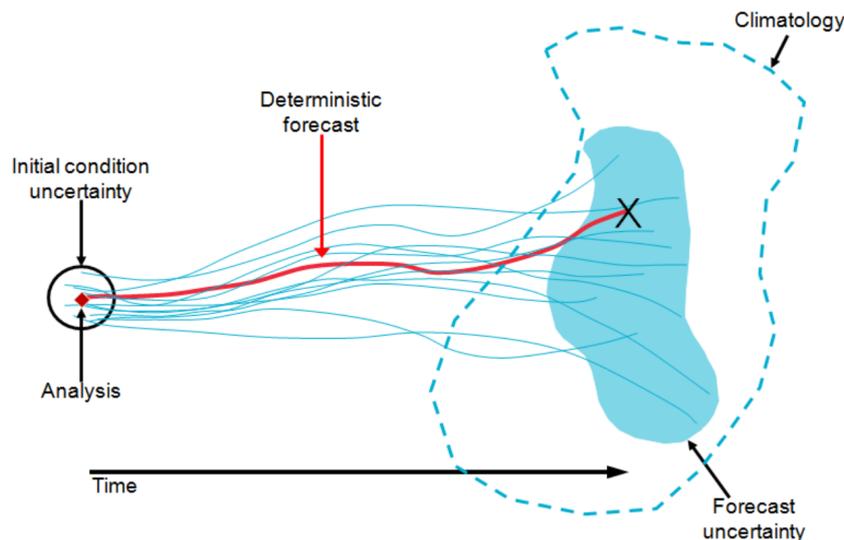


FIGURE 5 : Fonctionnement schématique d’une prévision d’ensemble, source : *A schematic illustration of an ensemble forecast* ©British Crown copyright, Met Office.

Chaque famille de prévision, par son horizon temporel, peut avoir un intérêt spécifique pour différents types d’utilisateurs. West et al. (2014) proposent une classification des utilisateurs de prévisions solaires. Le tableau 1 reprend quelques-unes de ces catégories

et présente pour chacune d'elle un exemple d'usage de prévision solaire en fonction de l'horizon temporel (cette liste d'usages n'est bien sûr pas exhaustive). Pour chaque cas, le tableau 1 propose aussi un article de référence détaillant l'usage spécifique de la prévision.

Acteur	Prévision infra-horaire	Prévision infra-journalière	Prévision J+1
Gestionnaire de réseau	Contrôler la stabilité du réseau <i>Orwig et al. (2012)</i>	Calculer une allocation dynamique des énergies de réserve <i>Bucksteeg et al. (2016)</i>	Dimensionner l'énergie de réserve nécessaire à la stabilité du réseau <i>Pierro et al. (2019)</i>
Producteur/trader	Valoriser la production par une vente sur le marché "intraday" <i>Aïd et al. (2016)</i>	Valoriser la production par une vente sur le marché "day-ahead" <i>Zwaenepoel et al. (2014)</i>	
Gestionnaire de micro-réseau	Assurer l'équilibre production/consommation du micro-réseau en gérant de manière optimale les ressources disponibles (batterie, autres sources d'énergie etc.) <i>Ramahatana et David (2019)</i>		

TABLEAU 1 : Différents usages possibles des prévisions solaires, selon le type d'acteur et l'horizon temporel de prévision.

Les prévisions les plus couramment utilisées opérationnellement sont les prévisions déterministes (aussi appelées prévision-point). Une prévision de cette nature fournit la prévision sous la forme d'une valeur unique. Or, dès les années 1960, Lorenz (1963) a iden-

tifié une catégorie de phénomènes caractérisés par leur extrême sensibilité aux conditions initiales. Li et Yorke (1975) les ont nommés phénomènes chaotiques; leur prédictibilité est intrinsèquement difficile, notamment parce que toute incertitude dans les conditions initiales peut entraîner des erreurs de prévision extrêmement importantes. À la suite de leur travail, la nature chaotique des phénomènes météorologiques a été démontrée, par exemple par Shen et al. (2021).

Il en résulte que les incertitudes associées aux prévisions des phénomènes météorologiques peuvent être dans certains cas importantes. Elles dépendent de nombreux paramètres comme les conditions climatologiques du site, la saison, l'heure de la journée ou les dynamiques atmosphériques, et l'utilisation d'une prévision déterministe ne permet pas d'en rendre compte. Par opposition à une prévision déterministe, une prévision probabiliste fournit une information supplémentaire sur l'incertitude associée à la prévision. Ces prévisions ont historiquement été utilisées dans certains domaines particuliers comme la macro-économie, le sport, la démographie, les assurances, etc. (Forrest et Simmons, 2000; Sims, 1993).

Compte-tenu de leur nature chaotique, le recours aux prévisions probabilistes pour les phénomènes météorologiques semble naturel. L'utilisation de ces prévisions pour la production d'énergie éolienne a débuté dans les années 1980 (Schueller et al., 1976). Certains travaux (Slougher et al., 2010; Roulston et al., 2003) ont souligné leur intérêt, en soulignant la valeur ajoutée non négligeable de l'information supplémentaire qu'elles portent.

L'utilisation des prévisions probabilistes pour la production d'énergie solaire est beaucoup plus récente, et encore largement minoritaire. Nous souhaitons dans ce travail démontrer qu'elles devraient pourtant être privilégiées, et que les informations relatives à l'incertitude constituent une aide à la prise de décision dans la gestion du réseau (pour le gestionnaire), dans la stratégie de vente d'électricité (pour les producteurs), ou encore dans la gestion des batteries dans les cas de micro-réseaux. C'est bien prioritairement à destination des acteurs industriels que la communauté scientifique propose ce type de prévision. Or, ces prévisions semblent au premier abord plus difficiles à appréhender et à manipuler, ce qui freine considérablement leur développement à grande échelle. Nous pensons qu'un travail illustrant leurs atouts pourrait participer à leur démocratisation. Par ailleurs, la prévision probabiliste appliquée à l'énergie solaire a été mise en avant comme un sujet à développer dans la Task 16 de l'IEA PVPS (2017-2023). Pour toutes ces raisons, ce travail de thèse a pour but principal de démontrer l'intérêt et la valeur des prévisions solaires probabilistes.

0.6 Contexte scientifique spécifique de la thèse : objectifs et démarche

Au vu des éléments exposés dans la partie 0.5, il nous semble que l'utilisation de prévisions probabilistes semble être une piste prometteuse en vue de faciliter l'intégration de l'énergie solaire au sein du mix électrique. Le but principal de ce travail est ainsi de déterminer si cette forme de prévision appliquée à la prévision solaire peut être à même d'apporter des bénéfices opérationnels aux acteurs de l'énergie. Il apparaît alors nécessaire de se tourner vers le travail de Murphy (1993), qui distingue différents types de bénéfices pouvant être attribués à une prévision, à savoir la consistance, la qualité et la valeur.

Un système de prévision météorologique dispose d'un ensemble d'informations qui l'amène à juger de la vraisemblance d'un évènement. La consistance d'une prévision se réfère au degré de correspondance entre la prévision finalement fournie et ce jugement interne. Comme ce dernier est impossible à connaître, il ne peut exister d'outils vérifiant la consistance.

La qualité d'un système de prévision est définie comme le degré de correspondance entre les prévisions et les observations. De nombreux outils de vérification de la qualité des prévisions, probabilistes ou déterministes, ont été proposés dans la littérature. Ils permettent de juger et de classer différentes prévisions via des critères mathématiques. Il ne faut pourtant pas oublier que ces prévisions ont pour but premier de fournir une aide aux acteurs industriels, et la qualité n'est pas reliée par définition aux bénéfices opérationnels qu'on cherche à mesurer dans ce travail.

La valeur d'une prévision est une propriété bien différente de sa qualité. C'est une mesure de l'utilité que la prévision fournira à son utilisateur. Certains travaux ont été consacrés à son étude (Perez et al., 2019; Roulston et al., 2003), mais elle reste relativement peu étudiée comparativement à la qualité. Nous pensons que c'est une lacune dans la littérature scientifique, car c'est justement à travers des démonstrations de la valeur ajoutée des prévisions probabilistes sur des cas d'études concrets que les acteurs industriels pourront s'appropriier ces prévisions et saisir leur intérêt. Nous souhaitons dans ce travail participer à convaincre les acteurs opérationnels ou industriels du secteur de l'électricité de la pertinence, pour eux, de l'utilisation de prévisions probabilistes. C'est pourquoi la valeur des prévisions est une notion qui sera longuement discutée dans ce travail de thèse.

Or, il n'est pas établi que des critères évaluant la qualité soient pertinents pour juger de la valeur des prévisions. Autrement dit, la corrélation entre qualité et valeur d'une prévision n'est pas certaine. Même si tel est le cas, il semble important de pouvoir sélectionner l'outil de vérification de la qualité le plus adapté parmi tous ceux qui existent, par exemple, celui qui reflète le mieux les coûts de l'utilisateur. Comme la littérature ayant trait à la valeur des prévisions est peu fournie, au contraire de celle étudiant leur qualité, nous souhaitons établir des liens entre qualité et valeur. S'ils peuvent être généralisés, l'établissement de tels liens permettrait à l'abondante littérature sur la qualité des prévisions d'apporter immédiatement des enseignements sur la valeur. Ces considérations amènent *in fine* à poser la problématique scientifique suivante :

Quels liens peut-on établir entre qualité et valeur d'une prévision solaire probabiliste ?

La réponse à cette problématique se fait à travers deux étapes distinctes. La première est dédiée à l'évaluation de la qualité de prévisions. Elle se décline en 2 sous-étapes à savoir : définir un cadre d'évaluation de ces prévisions permettant d'en étudier la qualité et mettre en évidence les principaux facteurs influant sur cette qualité. La seconde étape vise à présenter des méthodologies d'évaluation de la valeur. Compte-tenu de la définition de la valeur, ces méthodologies s'appuient sur une étude de cas. À l'issue de cette étape, nous tentons de conclure sur les liens possibles entre qualité et valeur des prévisions. La figure 6 résume la démarche de travail suivie.

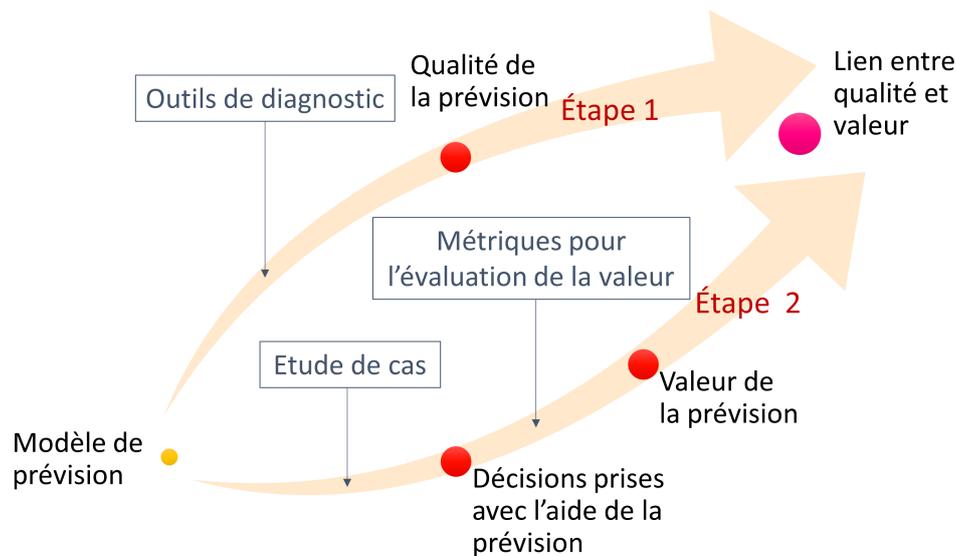


FIGURE 6 : Objectif et démarche du travail de mise en relation de la qualité et de la valeur des prévisions.

En plus de répondre à la problématique scientifique soulevée, un autre but de ce travail est d'illustrer l'intérêt des approches probabiliste et ensembliste. Les modèles de prévision pour lesquels la démarche présentée sur la figure 6 est appliquée ont donc été regroupés en quatre familles :

- A. Les prévisions déterministes construites uniquement à l'aide du membre de contrôle de la prévision d'ensemble ECMWF,
- B. Les prévisions déterministes issues de tous les membres d'une prévision d'ensemble,
- C. Les prévisions probabilistes construites à partir du membre de contrôle, afin de mesurer les bénéfices de l'approche probabiliste, par comparaison avec les prévisions de type "A",
- D. Les prévisions probabilistes issues de tous les membres d'une prévision d'ensemble, afin de mesurer les bénéfices de l'approche ensembliste (par comparaison avec les prévisions de type "C") et de l'approche probabiliste (par comparaison avec les prévisions de type "B").

Les bénéfices retirés par les approches probabiliste et ensembliste sont alors mesurés à l'aune des différences obtenues en terme de qualité et de valeur à l'issue des étapes 1 et 2. Afin d'utiliser les prévisions d'ensemble dans leur domaine de prédilection, nous travaillons uniquement sur des prévisions $J+1$.

0.7 Structure du manuscrit

Le manuscrit est structuré de façon à suivre les étapes de travail évoquées dans la partie 0.6. L'organisation du manuscrit est représentée schématiquement par la figure 7.

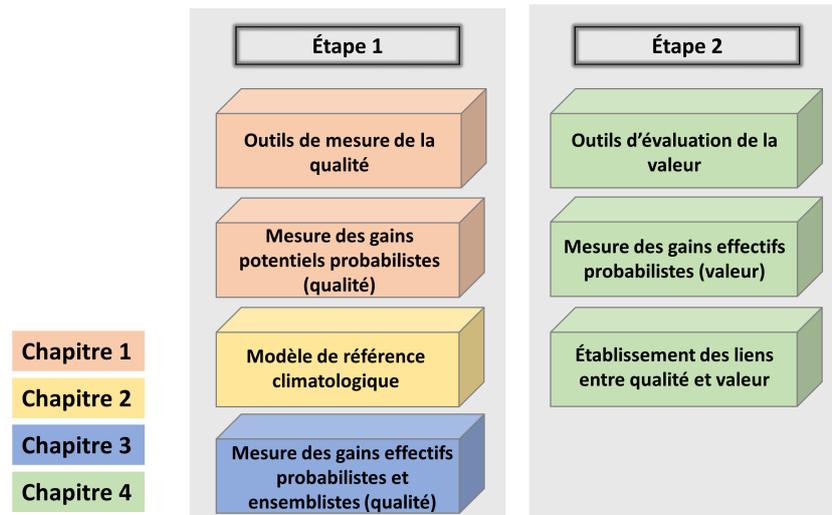


FIGURE 7 : Organisation schématique du manuscrit.

L'étape 1 est divisée en 3 chapitres distincts. Pour débiter, le chapitre 1 propose une étude bibliographique, qui recense certains outils de vérification de la qualité des prévisions solaires probabilistes. En particulier, elle a pour but de rechercher dans la littérature des métriques permettant d'évaluer à la fois des prévisions probabilistes et déterministes. De cette façon, on peut espérer comprendre si elles permettent d'envisager un gain de qualité pour l'approche probabiliste et évaluer le cas échéant ce gain maximum potentiel. C'est pourquoi au-delà de cet état de l'art, nous proposons une nouvelle démarche permettant la comparaison des métriques probabilistes, et nous l'utilisons pour quantifier les gains potentiels apportés par l'utilisation de prévisions probabilistes.

Cependant, les bonnes pratiques à adopter pour mesurer objectivement et correctement la qualité des prévisions ne se réduisent pas seulement à l'utilisation de métriques. Elles doivent aussi comprendre des modèles de référence permettant de quantifier par comparaison la performance des modèles de prévision proposés. Le cadre d'évaluation doit aussi s'appuyer sur des données d'éclaircissement énergétique publiquement accessibles, documentées et validées selon une procédure définie. Cette recherche bibliographique est donc suivie dans le chapitre 2 par la proposition d'un nouveau modèle de référence ("benchmark model") qui permet la comparaison de prévisions issues de différents contextes. Comme des données de rayonnement sont déjà disponibles publiquement grâce notamment au réseau BSRN ("Baseline Surface Radiation Network", <https://bsrn.awi.de/>), ce deuxième chapitre complète le premier afin de proposer un cadre général d'évaluation des prévisions probabilistes.

Pour clôturer le travail sur la qualité des prévisions, le chapitre 3 cherche à mesurer sur des cas pratiques les gains de qualité réellement apportés par les approches probabilistes et ensemblistes. De plus, un travail de recherche est effectué sur les facteurs pouvant impacter la qualité des prévisions. Ce faisant, nous proposons une nouvelle méthodologie utilisant les moments statistiques des distributions pour évaluer ces gains.

L'étape 2 qui se concentre sur la valeur des prévisions est traitée dans le chapitre 4. Nous proposons un état de l'art des différentes méthodes existantes d'évaluation de la valeur. Nous établissons aussi une méthodologie permettant d'évaluer la valeur sur des cas d'étude. Enfin, nous illustrons l'utilisation de cette méthodologie à travers une étude de cas. La question des liens entre qualité et valeur est également discutée dans ce chapitre.

Chapitre 1

Évaluation de la qualité des prévisions probabilistes

Résumé

Ce chapitre met en lumière les différences d'approche entre la prévision probabiliste et la prévision déterministe. Il explicite aussi les différentes formes que peut prendre une prévision probabiliste. L'accent est ensuite mis sur les prévisions probabilistes. La littérature scientifique a discerné plusieurs attributs propres aux prévisions probabilistes. Ils sont présentés dans une première partie. Dans un second temps, un état de l'art sur les outils de vérification de la qualité des prévisions probabilistes est présenté. Ces derniers peuvent prendre la forme d'un score quantifiant la qualité ou d'un outil visuel vérifiant un attribut spécifique de la prévision. À l'issue de cette revue bibliographique, une carte des scores propres probabilistes est établie. Elle propose de séparer les scores selon leur famille et la forme de prévision probabiliste concernée.

Les scores identifiés dans l'étude bibliographique sont ensuite utilisés pour tenter de déterminer le gain de qualité potentiel associé à l'utilisation des prévisions probabilistes. Pour aider à mesurer ce gain pour les prévisions de variables binaires, un nouvel outil de représentation des scores binaires est élaboré : la matrice d'espérance. On parvient à déterminer que le choix du score et son caractère strictement propre ou non sont cruciaux dans la mesure de ce gain. Cette partie conduit aussi aux sélections préférentielles du *CRPS* ("*Continuous Ranked Probability Score*") et du *quantile score* pour la suite du travail. À la fois pour appréhender le gain potentiel de l'approche probabiliste et pour illustrer sur un cas concret les attributs de ces prévisions probabilistes, une étude analysant l'évolution de ces attributs et du *CRPS* au cours de l'amélioration continue d'une prévision est présentée.

Sommaire

1.1	Introduction	17
1.2	Notions préliminaires	19
1.3	Les attributs relatifs à la qualité d'une prévision probabiliste	20
1.3.1	La fiabilité	20
1.3.2	L'incertitude	21
1.3.3	La résolution	21

1.3.4	La finesse	21
1.3.5	La calibration	21
1.4	Outils d'évaluation de la qualité	22
1.4.1	Propriétés des scores	22
1.4.2	Revue bibliographique des scores probabilistes	25
1.4.3	Carte des scores propres	30
1.4.4	Les scores de compétences ou "skill scores"	31
1.4.5	Outils graphiques	32
1.5	Comparaison des qualités des prévisions déterministes et probabilistes	37
1.5.1	Les matrices d'espérance	38
1.5.2	Gains en qualité de l'approche probabiliste pour les scores de variables binaires	41
1.5.3	Gains de l'approche probabiliste pour les scores de variables continues	44
1.5.4	D'une mauvaise prévision à la prévision parfaite	47
1.6	Conclusions	51
	Annexes	53
1.A	Démonstration du caractère propre des scores binaires de type "fonction puissance"	53
1.B	Décomposition du <i>Brier score</i>	54
1.C	Équivalences du <i>Brier score</i> , du <i>score propre linéaire</i> et du <i>Winkler score</i>	56

1.1 Introduction

La croissance de la production d'électricité issue de l'énergie solaire intermittente complique les opérations de “dispatching” des différents acteurs des réseaux électriques. Pour remédier à ce problème, la communauté scientifique étudie comment prévoir le plus précisément possible la production photovoltaïque (PV). Celle-ci est directement liée au rayonnement solaire horizontal global (ou “GHI”).

Il est possible de discerner trois formes différentes d'évènements auxquels peuvent être associées des prévisions :

- Des évènements (ou variables) binaires : l'évènement expliqué par la prévision se déroule ou non. De telles prévisions peuvent répondre par exemple aux questions : Pleuvra-t-il demain ? L'ensoleillement va-t-il excéder 500 W/m^2 ? etc.
- Des variables multi-catégories, pour lesquelles l'évènement peut être classé dans une catégorie parmi un nombre fini de catégories (par exemple : l'éclairement énergétique sera-t-il inférieur à 200 W/m^2 , compris entre 200 et 500 W/m^2 ou supérieur à 500 W/m^2 ?). Ce type de prévision peut être compris comme une extension des prévisions binaires, lorsque le nombre de catégories est supérieur à 2.
- Des variables continues, où l'évènement sera caractérisé par une valeur mesurée parmi un nombre infini de valeurs possibles, par exemple : Quelle sera la valeur du GHI, en W/m^2 ?

Les prévisions les plus couramment utilisées sont des prévisions déterministes, qui fournissent une valeur unique de prévision. La recherche de nouvelles techniques de prévision est le sujet d'abondantes recherches dans la littérature (Chen et al., 2013 ; Azimi et al., 2016 ; Kumler et al., 2019 ; Perez et al., 2007). On définit la qualité d'une prévision comme sa précision (Thornes et Stephenson, 2001), ou sa correspondance avec les observations (Murphy, 1993). Casati et al. (2008) rappellent l'importance de disposer d'un cadre objectif d'évaluation de la qualité des prévisions. Ce cadre permet d'effectuer des comparaisons justes, de critiquer une prévision, de classer entre elles différentes prévisions, ou encore de sélectionner objectivement une prévision parmi plusieurs. Régulièrement, certains travaux proposent de nouveaux outils graphiques (Roebber, 2009) ou de nouvelles métriques (Vallance et al., 2017) d'évaluation de la qualité pour enrichir ce cadre. D'autres (Yang et al., 2020b) se consacrent à synthétiser certains de ces outils.

Cependant, pour rendre compte de l'incertitude associée aux prévisions, les prévisions probabilistes suscitent un intérêt croissant au sein de la communauté scientifique. Ce type de prévision, encore peu utilisé opérationnellement, est adapté à des évènements peu prédictibles, souvent chaotiques. Il fournit une quantité d'information plus importante que la prévision déterministe, notamment à propos de l'incertitude associée à la prévision.

Les approches associées aux prévisions déterministes et probabilistes sont sensiblement différentes. Les prévisions déterministes présument que le phénomène est certes encore inconnu, mais déterminé. Choisir une prévision probabiliste concède au contraire, du point de vue du prévisionniste et au moment de la prévision, que l'évènement n'est pas encore complètement déterminé. Ce dernier est alors considéré comme une variable aléatoire qui suit une loi de probabilité, qu'on appellera par la suite “loi de probabilité réelle” caractérisée par sa fonction de répartition \mathcal{Q} . La prévision probabiliste n'a plus pour objectif de prédire l'observation finale, mais la loi de probabilité de l'évènement. C'est

pourquoi les prévisions probabilistes fournissent non plus une valeur unique, mais une prévision sous la forme d'une loi de probabilité ayant une fonction de répartition $\hat{\mathcal{F}}$, qui a pour but d'être la plus proche possible de la fonction de répartition réelle \mathcal{Q} . La différence entre une prévision déterministe et une prévision probabiliste est illustrée sur la figure 1.1

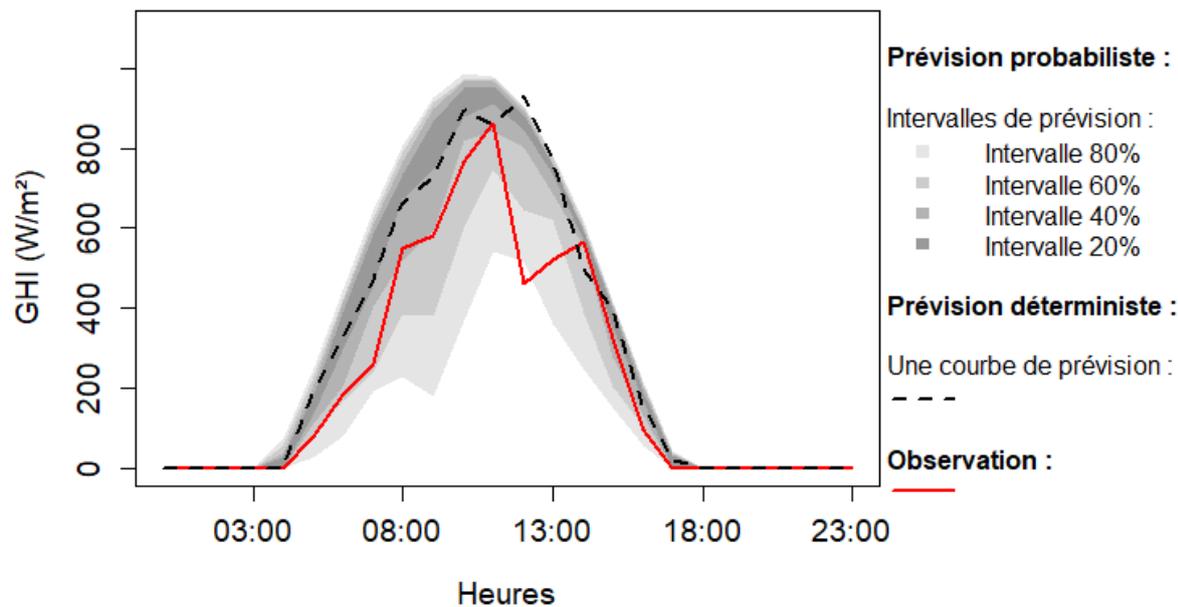


FIGURE 1.1 : Comparaison entre une prévision déterministe et une prévision probabiliste sous la forme d'intervalles de probabilités.

Une prévision probabiliste peut prendre différentes formes, parmi lesquelles on peut citer :

- Une prévision sous la forme d'une fonction de probabilité,
- Une prévision dite "prévision d'ensemble" qui est un ensemble de valeurs représentant chacune un scénario possible,
- Une prévision dite "prévision quantile", qui cherche à estimer de manière discrète la valeur du quantile de la distribution pour un niveau de probabilité τ donné.

Pour générer des prévisions probabilistes, il existe une multitude de techniques différentes. Elles diffèrent par leurs hypothèses sous-jacentes (méthodes paramétriques ou non), les données d'entrée des modèles (ensoleillement énergétique horizontal, température, humidité, etc.), et leur type (données exogènes ou endogènes). De la même façon que pour les prévisions déterministes, il est nécessaire de pouvoir mesurer objectivement les performances des prévisions ainsi créées, de comparer et opérer des tris entre ces différents modèles de prévision. Or, les outils de vérification des prévisions déterministes ne sont en général pas applicables aux prévisions probabilistes.

La communauté scientifique a alors développé un cadre indépendant d'évaluation des prévisions probabilistes. Poursuivant le travail initié par Brier (1950), le développement

de ce cadre s’est fait ces dernières décennies via la proposition de nombreuses métriques différentes, de plusieurs outils de vérification graphiques, mais aussi via le développement d’une structure théorique permettant d’asseoir les outils de vérification proposés. En particulier, cette recherche théorique a mis en lumière l’importance d’utiliser des “scores propres”¹ pour évaluer des prévisions probabilistes (Gneiting et Raftery, 2007 ; Bröcker, 2009b ; Bröcker, 2009a), c’est-à-dire des scores qui, de par leur définition, incitent le prévisionniste à fournir des prévisions honnêtes approchant le plus possible la loi de probabilité réelle de la variable aléatoire cible.

Ce premier chapitre a pour but d’évaluer le gain de qualité potentiel induit par l’utilisation de prévisions probabilistes, c’est-à-dire le gain maximal qu’il est possible d’obtenir grâce à ce type de prévision. Pour cela, il est nécessaire d’identifier des outils de vérification de la qualité pouvant comparer des prévisions déterministes ou probabilistes. Comme les outils de vérification déterministes ne sont pas adaptables aux prévisions probabilistes, on recherche ici si des outils de vérification probabilistes peuvent accomplir cet objectif et fournir des éléments de comparaison. Si c’est le cas, on tentera d’analyser ces outils et de mesurer le gain potentiel que ceux-ci accordent à l’approche probabiliste. On recherchera aussi quels outils sont les plus adaptés pour répondre au but général de cette thèse qui est d’établir des liens entre qualité et valeur. En particulier, on éliminera les scores pour lesquels les gains de l’approche probabiliste sont infinis ou non définis, car de tels gains ne pourront pas être mis en relation avec les gains en valeur, qui sont nécessairement finis.

Ce chapitre est structuré comme suit. La partie 1.2 rappelle quelques notions préliminaires indispensables. La partie 1.3 présente différents attributs qu’on peut lier à une prévision probabiliste. Un état de l’art portant sur les outils de vérification de la qualité des prévisions probabilistes est présenté dans la partie 1.4. Ensuite, la partie 1.5 recherche si les outils identifiés peuvent nous renseigner sur les gains de qualité induits pas l’approche probabiliste. Dans ce but, un nouvel outil de visualisation des scores probabilistes est proposé : la matrice d’espérance.

1.2 Notions préliminaires

On appellera “système de prévision” un ensemble de couples prévision/observation, où toutes les prévisions sont produites par une même méthode.

Une variable binaire est un événement dont les réalisations possibles sont 0 (l’évènement ne s’est pas réalisé) ou 1 (l’évènement s’est réalisé). Ces deux réalisations sont respectivement prédites par les probabilités \hat{F}_0 et \hat{F}_1 . Cet évènement peut être le dépassement d’un seuil par une variable continue. Il convient de rappeler que pour une prévision binaire, \hat{F}_0 et \hat{F}_1 sont liées par $\hat{F}_0 + \hat{F}_1 = 1$. Par ailleurs, on utilisera dans ce chapitre la notation \hat{F}_O pour désigner la probabilité assignée par la prévision à la catégorie finalement observée, c’est-à-dire que :

$$\hat{F}_O = \begin{cases} \hat{F}_0 & \text{si } O = 0, \\ \hat{F}_1 & \text{si } O = 1. \end{cases} \quad (1.1)$$

De même, pour une prévision multi-catégories, on a $\sum_{i=1}^I \hat{F}_i = 1$. On étend la notation \hat{F}_O à ce type de prévision.

¹Le terme consacré dans la littérature scientifique anglophone est “proper score” qui peut se traduire littéralement par “score correct”, ce qui sous-entend que les autres scores doivent être évités. Cependant, dans la littérature scientifique francophone, la traduction “score propre” s’est imposée par l’usage.

La prévision probabiliste d’une variable continue telle que le GHI peut être donnée sous la forme d’une fonction de répartition (ou fonction “CDF” pour “Cumulative Distribution Function”) qu’on nommera $\hat{\mathcal{F}}(X)$. Une telle fonction est croissante et telle que $\min_{\mathbb{R}} \hat{\mathcal{F}} = 0$ et $\max_{\mathbb{R}} \hat{\mathcal{F}} = 1$. Elle peut aussi être donnée de manière équivalente par sa dérivée, une fonction de densité de probabilité (ou fonction “PDF” pour “Probability Density Function”) qu’on notera $\hat{f}(X)$, et qui suivra nécessairement $\int_{\mathbb{R}} \hat{f}(X) dX = 1$. De plus, pour les variables continues, la fonction CDF associée à une mesure O et notée \mathcal{O} est dans ce travail définie par² :

$$\mathcal{O}(x) = H(x - O), \quad (1.2)$$

où H est la fonction de Heaviside.

La climatologie se réfère à la moyenne des observations sur le long terme. Elle peut être estimée sur une période d’apprentissage comportant N_A observations par :

- $\bar{O} = \frac{1}{N_A} \sum_{n=1}^{N_A} O_n$ pour une variable binaire,
- $\bar{\mathcal{Q}} = \frac{1}{N_A} \sum_{n=1}^{N_A} \mathcal{O}_n(x)$ pour une variable continue.

1.3 Les attributs relatifs à la qualité d’une prévision probabiliste

Murphy (1993) met en évidence un certain nombre de propriétés que l’on peut attribuer à un système de prévision probabiliste. Ces propriétés sont aujourd’hui souvent étudiées pour effectuer un diagnostic fin des caractéristiques d’un système de prévision (Lauret et al., 2019; Lauret et al., 2017), ou pour en comprendre les points forts ou faibles. Certaines de ces propriétés, comme la discrimination (Weigel et Mason, 2011), sont obtenues à partir de la “likelihood-base rate factorization” d’un ensemble de couples prévisions/observations, c’est-à-dire l’étude des distributions de probabilité de la prévision conditionnées par l’observation $p(\hat{F}|O)$ (Wilks, 2014). Dans cette partie, celles-ci sont mises de côté et les propriétés présentées sont issues de l’autre factorisation possible appelée “calibration-refinement factorization” qui consiste en l’étude des distributions de probabilité de l’observation conditionnées par la prévision $p(O|\hat{F})$. Cinq de ces propriétés parmi les plus couramment évaluées sont listées ci-dessous.

1.3.1 La fiabilité

La fiabilité d’un système de prévision est un indicateur de la consistance statistique de la prévision. Elle évalue la capacité d’un système à fournir des probabilités de prévision proches des probabilités réelles (c’est-à-dire des fréquences réellement observées pour un grand nombre de couples prévision/observation). Une mesure de la fiabilité est donc une

²Il est possible de définir différemment \mathcal{O} dans des cas où une incertitude porte aussi sur l’observation.

réponse à la question : “Les probabilités indiquées par le système correspondent-elles aux fréquences réellement observées à long terme?”

1.3.2 L’incertitude

L’incertitude est un indicateur de variabilité de la variable cible. Elle est indépendante de la prévision, et est uniquement due à la distribution des observations. Autrement dit, elle ne peut pas être diminuée par le travail du prévisionniste : elle est “incompressible”. Elle peut être comprise comme un indice sur la difficulté intrinsèque de prévision de la variable cible.

1.3.3 La résolution

La résolution est la capacité d’un système de prévision à trier et regrouper correctement les différents cas de prévision, c’est-à-dire à créer des groupes d’observations dont la moyenne est éloignée de la moyenne climatologique. La résolution d’un système de prévision est élevée lorsque les distributions des observations pour diverses situations de prévision sont distinctes de la distribution climatologique.

1.3.4 La finesse

La finesse (ou “sharpness”) d’une prévision n’est définie que pour une prévision d’ensemble ou une prévision sous forme de fonction de probabilité. Elle quantifie la capacité d’un système à fournir des prévisions fines, c’est-à-dire mathématiquement parlant, avec un faible écart-type. Elle est en général déterminée par l’écart entre deux niveaux de probabilité de la fonction de prévision CDF. La finesse est un concept assez proche de la résolution. Cependant, contrairement à celle-ci, elle est un attribut de la prévision uniquement. L’obtention d’une prévision la plus fine possible est souhaitable, tant que l’accroissement de cette finesse ne s’effectue pas au détriment de la fiabilité (Gneiting et al., 2007).

1.3.5 La calibration

Les notions de calibration et de fiabilité sont très proches, cependant, elles ne sont pas identiques, notamment car la fiabilité est toujours définie conditionnellement à la prévision. Il existe plusieurs modes de calibration définis dans différentes études (Tsyplakov, 2013; Ben Bouallegue, 2017; Strähl et Ziegel, 2017; Gneiting et al., 2007). Les différents modes et les terminologies associées ne sont pas toujours consistantes entre ces études. Néanmoins, nous présentons ici 3 modes de calibration dont les définitions sont communément acceptées.

- La calibration *marginale* est vérifiée si la fonction CDF de prévision moyenne est asymptotiquement égale à la distribution des observations. Autrement dit, la calibration marginale vérifie l’égalité entre les climatologies observées et mesurées, ce qui revient formellement à vérifier si :

$$\lim_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{n=1}^N \hat{\mathcal{F}}_n \right) = \bar{\mathcal{Q}}. \quad (1.3)$$

- La calibration *probabiliste* vérifie si une probabilité donnée par le système de prévision est proche de la probabilité réelle. Une prévision est probabilistiquement calibrée si :

$$\frac{1}{N} \sum_{n=1}^N \mathcal{O}_n \circ \hat{\mathcal{F}}_n^{-1}(\tau) = \tau, \quad (1.4)$$

pour tout $\tau \in [0;1]$. La calibration probabiliste et la fiabilité sont des concepts similaires.

- La calibration *conditionnelle* vérifie la consistance statistique en fonction d'un paramètre extérieur. En général, les prévisions sont dépendantes de certains paramètres explicatifs (paramètres temporels, prévision d'ensemble, etc.). Tsyplakov (2013) recommande de vérifier la calibration par rapport à ces paramètres, c'est-à-dire de vérifier la consistance statistique des prévisions et des observations conditionnellement aux valeurs des paramètres explicatifs. Pour vérifier cette calibration, il faut vérifier les consistances marginales et probabilistes pour chaque valeur différente du paramètre explicatif.

1.4 Outils d'évaluation de la qualité

Le but de tout système de prévision est de fournir une information qui soit la plus précise possible. Pour évaluer cette précision, appelée qualité, il est nécessaire de disposer d'outils de vérification des prévisions. Pour évaluer les prévisions probabilistes, la littérature scientifique a proposé des outils spécifiques. Bien que certains outils de vérification des prévisions déterministes puissent être interprétés du point de vue probabiliste, en considérant que la variable cible X suit une loi de CDF \mathcal{Q} inconnue, ils ne sont pas utilisables pour évaluer une prévision probabiliste. Dans cette partie, nous présentons un état de l'art des outils de vérification des prévisions probabilistes. Ces outils peuvent être de deux natures différentes : des métriques, aussi appelées scores, qui permettent une évaluation quantitative de la qualité ou des outils graphiques. Le choix de l'outil dépend du type de prévision probabiliste (variable binaire ou continue, prévision de quantiles, etc.).

Dans une première partie 1.4.1, certaines propriétés importantes des scores probabilistes sont étudiées. Cette présentation est suivie par celles de plusieurs familles de scores probabilistes : les scores quadratiques, les scores linéaires, les scores pseudosphériques, les scores logarithmiques et les autres scores.

1.4.1 Propriétés des scores

Score et mesure d'erreur

Le score obtenu par un système de prévision est toujours calculé en moyennant sur une période de test les erreurs de chacun des couples prévision/observation. Certaines recherches nomment "score" la mesure de l'erreur pour un couple alors que d'autres nomment score la moyenne finalement effectuée caractérisant la performance globale du système de prévision. Dans ce travail, nous avons choisi d'appeler "mesure d'erreur associée à un score" ou " S_n " la mesure de l'erreur associée à un couple $(\hat{\mathcal{F}}_n, \mathcal{O}_n)$ et "score \bar{S} " la moyenne

finalemeut obtenue. Avec cette notation, on a donc :

$$\bar{S} = \frac{1}{N} \sum_{n=1}^N S_n(\hat{\mathcal{F}}_n, \mathcal{O}_n). \quad (1.5)$$

Cependant, pour alléger les notations, on fera référence à un score moyen \bar{S} par la notation simplifiée “ S ”.

Orientation et positivité

Un score est positif si sa définition mathématique implique qu’il ne puisse prendre que des valeurs positives. Un score négativement orienté est tel qu’une meilleure qualité est indiquée par un score plus faible (c’est le cas de tous les scores décrits dans ce chapitre, sauf mention contraire). Ainsi, un score positif et négativement orienté est tel que la meilleure valeur possible est 0.

Localité

Un score est dit “local” lorsque seule la probabilité de la PDF de prévision \hat{f} prise au niveau de l’observation O , soit $\hat{f}(O)$, est utilisée pour l’évaluation. Au contraire, il n’est pas local si plus d’informations issues de la prévision sont prises en compte dans le calcul du score. Autrement dit, un score est local si la mesure de son erreur peut être écrite sous la forme suivante (Bröcker et Smith, 2007a) :

$$S_n(\hat{f}, O) = S_n(\hat{f}(x = O), O). \quad (1.6)$$

S’il peut sembler plus naturel d’utiliser des score locaux, Bröcker et Smith (2007a) montrent que l’utilisation de scores non-locaux est préférable dans les cas où l’observation est elle-même mesurée avec un certain niveau d’incertitude.

Notion de score propre

Un score S est une mesure mathématique de la qualité d’un système de prévision, c’est-à-dire d’un ensemble de couples prévision/observation. Une prévision probabiliste est utilisée quand on considère la variable cible comme une variable aléatoire, qui possède donc par définition une loi de probabilité \mathcal{Q} . La mesure d’erreur associée à un score S est donc une fonction qui associe à une prévision probabiliste $\hat{\mathcal{F}}$ et à une fonction observation \mathcal{O} une valeur $S_n(\hat{\mathcal{F}}, \mathcal{O})$ qui quantifie l’erreur. \mathcal{O} étant la réalisation de la variable aléatoire X qui suit la loi de CDF \mathcal{Q} , on définit $\mathbb{E}[S_n(\hat{\mathcal{F}}, \mathcal{Q})]$, l’espérance pour S_n d’une prévision $\hat{\mathcal{F}}$ sachant que la distribution réelle de X est \mathcal{Q} . Appelons \mathcal{E} l’ensemble de toutes les prévisions possibles. Un score S négativement orienté est dit “score propre” si :

$$\mathbb{E} \left[S_n(\hat{\mathcal{F}}, \mathcal{Q}) \right] \geq \mathbb{E} \left[S_n(\mathcal{Q}, \mathcal{Q}) \right], \quad (1.7)$$

pour toutes les prévisions $\hat{\mathcal{F}} \in \mathcal{E}$. De plus, un score est dit strictement propre si :

$$\mathbb{E} \left[S_n(\hat{\mathcal{F}}, \mathcal{Q}) \right] = \mathbb{E} \left[S_n(\mathcal{Q}, \mathcal{Q}) \right] \iff \hat{\mathcal{F}} = \mathcal{Q}. \quad (1.8)$$

En pratique, ces espérances ne peuvent pas être calculées car la loi de probabilité réelle \mathcal{Q} reste inconnue et seule O est mesurée.

Décomposition d'un score propre

Étant donné l'équation 1.7, utiliser un score propre négativement orienté permet de définir pour toute prévision $\hat{\mathcal{F}}$ une quantité $div_S(\hat{\mathcal{F}}, \mathcal{Q})$ qui est telle que :

$$\mathbb{E} \left[S_n(\hat{\mathcal{F}}, \mathcal{Q}) \right] = \mathbb{E} \left[S_n(\mathcal{Q}, \mathcal{Q}) \right] + div_S(\hat{\mathcal{F}}, \mathcal{Q}) \text{ avec } div_S(\hat{\mathcal{F}}, \mathcal{Q}) \geq 0. \quad (1.9)$$

$div_S(\hat{\mathcal{F}}, \mathcal{Q})$ est appelée divergence de S entre $\hat{\mathcal{F}}$ et \mathcal{Q} , et $\mathbb{E}[S(Q, Q)]$ est appelé entropie de la distribution \mathcal{Q} pour S (Bröcker, 2009b). On définit aussi la fonction entropie \mathcal{H}_S d'un score S pour l'ensemble des $\hat{\mathcal{F}} \in \mathcal{E}$ par :

$$\mathcal{H}_S(\hat{\mathcal{F}}) = \mathbb{E}[S_n(\hat{\mathcal{F}}, \hat{\mathcal{F}})]. \quad (1.10)$$

En considérant la distribution climatologique $\bar{\mathcal{Q}}$, on peut développer :

$$\mathbb{E} \left[S_n(\mathcal{Q}, \mathcal{Q}) \right] = \mathbb{E} \left[S_n(\bar{\mathcal{Q}}, \mathcal{Q}) \right] - div_S(\bar{\mathcal{Q}}, \mathcal{Q}), \quad (1.11)$$

puis, en combinant les équations 1.9 et 1.11 :

$$\mathbb{E} \left[S_n(\hat{\mathcal{F}}, \mathcal{Q}) \right] = \mathbb{E} \left[S_n(\bar{\mathcal{Q}}, \mathcal{Q}) \right] - div_S(\bar{\mathcal{Q}}, \mathcal{Q}) + div_S(\hat{\mathcal{F}}, \mathcal{Q}). \quad (1.12)$$

On peut alors définir :

$$REL = div_S(\hat{\mathcal{F}}, \mathcal{Q}), \quad (1.13)$$

$$RES = div_S(\bar{\mathcal{Q}}, \mathcal{Q}), \quad (1.14)$$

$$UNC = \mathbb{E} \left[S_n(\bar{\mathcal{Q}}, \mathcal{Q}) \right]. \quad (1.15)$$

L'intérêt de cette décomposition est qu'elle est toujours théoriquement possible, dès lors qu'un score propre est utilisé. On obtient ainsi la décomposition canonique d'un score propre :

$$\mathbb{E} \left[S_n(\hat{\mathcal{F}}, \mathcal{Q}) \right] = REL - RES + UNC. \quad (1.16)$$

REL est appelée la fiabilité, UNC l'incertitude et RES la résolution. Les conséquences respectives des équations 1.9 et 1.11 sont que la fiabilité est positive et négativement orientée, et la résolution est comprise entre 0 et UNC et est positivement orientée. Comme \mathcal{Q} reste inconnue, les valeurs réelles de la fiabilité, de la résolution et de l'incertitude ne peuvent théoriquement pas être calculées. Cependant, pour un système de prévision incluant un nombre important de couples prévision/observation, il est possible de faire une estimation de ces quantités moyennées. Cette estimation est faite en agrégeant ensemble les observations selon les valeurs de la prévision. On suppose que pour chaque prévision, \mathcal{Q} est égale à la distribution des observations ainsi agrégées. C'est pourquoi les scores sont calculés en pratique pour un nombre important de couples prévision/observation (Candille et Talagrand, 2005). Cette procédure permet de passer de la décomposition

théorique du score S_n d'un couple prévision/observation à la décomposition de S et pour certains scores propres de calculer les valeurs de fiabilité, résolution et incertitude d'un système de prévision.

Étudier la décomposition du score d'un système de prévision permet d'obtenir des informations supplémentaires sur le type d'erreur de la prévision. Les interprétations de ces différents composants ont été détaillées dans la partie 1.3. En particulier, on peut noter que $RES = U$ montre que le système de prévision parvient à capter toute la variabilité de la variable cible, alors que $RES = 0$ montre que la prévision donnée par le système est la même quelle que soit l'observation, ce qui est le cas par exemple de la prévision climatologique (cf. chapitre 2).

Avant l'utilisation d'une métrique, l'importance de vérifier que le score utilisé est propre a été souligné de nombreuses fois Bröcker et Smith (2007a), et ce d'autant plus qu'il existe encore dans la littérature des exemples d'utilisation de scores non propres, ce qui nécessite des vérifications ou des corrections (Pinson et Tastu, 2014).

Scores équivalents

Pour suivre la terminologie proposée par Dawid et al. (2012) et Dawid (2006), Des scores S^1 et S^2 sont qualifiés d'équivalents si leurs mesures d'erreur sont reliées par une relation de la forme :

$$S_n^2 = a \times S_n^1 + b, \quad (1.17)$$

où a et b sont des constantes avec $a > 0$.

Les divergences de scores équivalents sont égales. Des scores équivalents suggèrent des conclusions identiques, et classeront toujours des prévisions de la même façon.

1.4.2 Revue bibliographique des scores probabilistes

Les scores quadratiques

Le *Brier score* a été le premier outil d'évaluation proposé pour les prévisions probabilistes. Il a été créé par Brier (1950) puis s'est démocratisé grâce au travail de Murphy (1969). Il s'applique à la prévision d'un évènement binaire et est défini par :

$$BS = \frac{1}{N} \sum_{n=1}^N (\hat{F}_{1,n} - O_n)^2. \quad (1.18)$$

Il vient de cette définition que le *Brier score* est un score compris entre 0 et 1 et négativement orienté.

La démonstration que le *Brier score* est propre est proposée dans l'Annexe 1.A et le calcul de sa décomposition en fiabilité, résolution et incertitude est disponible dans plusieurs travaux (Young, 2010 ; Candille et Talagrand, 2005 ; Ben Bouallegue, 2017).

Il existe une extension du *Brier score* aux prévisions multi-catégories appelée le "*Ranked Probability Score*" ou *RPS*. On peut considérer qu'une prévision multi-catégories est un ensemble de prévisions binaires, indexées par exemple par i , et estimant toutes la probabilité de dépassement d'un seuil. Sommer les *Brier score* de toutes ces prévisions revient à effectuer le calcul suivant :

$$RPS = \frac{1}{N} \sum_{n=1}^N \left[\sum_{i=1}^I (\hat{F}_{i,n} - O_{i,n})^2 \right], \quad (1.19)$$

qui définit le *RPS*. On peut remarquer le lien fort qui existe entre *Brier score* et *RPS* en constatant que le *RPS* est égal au *Brier score* dans le cas où $I = 2$.

Le *CRPS* (pour “*Continuous Ranked Probability Score*”) est un score propre couramment utilisé et adapté à l’évaluation de prévisions de variables continues telles que le GHI. Sa définition est donnée par :

$$CRPS = \frac{1}{N} \sum_{n=1}^N \int_{-\infty}^{\infty} \left[\hat{\mathcal{F}}_n(x) - \mathcal{O}_n(x) \right]^2 dx. \quad (1.20)$$

Machete (2013) prouve que le *CRPS* est un score propre. Le *CRPS* a aussi la propriété de pouvoir être défini comme l’intégrale du *Brier score* sur l’ensemble des valeurs prises par la variable cible :

$$CRPS = \int_{-\infty}^{\infty} BS(x) dx. \quad (1.21)$$

En pratique, dans le cas particulier de la prévision du GHI, il n’est pas nécessaire d’effectuer l’intégration entre $-\infty$ et ∞ , car les valeurs possibles du GHI sont contraintes pour des raisons physiques. Long et Shi (2008) proposent les bornes $Clim_{min} = -4 \text{ W/m}^2$ et $Clim_{max} = 1.5 \times Sa \times \mu_0^{1.2} + 100 \text{ W/m}^2$ où Sa est la constante solaire et μ_0 est le cosinus de l’angle solaire zénital. La décomposition du *CRPS* découle naturellement de la décomposition du *Brier score* et ses composants *REL_{CRPS}*, *UNC_{CRPS}* et *RES_{CRPS}* peuvent être exprimés en fonction des composants du *Brier score* *REL_{BS}*, *UNC_{BS}* et *RES_{BS}* :

$$REL_{CRPS} = \int_{Clim_{min}}^{Clim_{max}} REL_{BS}(x) dx, \quad (1.22)$$

$$RES_{CRPS} = \int_{Clim_{min}}^{Clim_{max}} RES_{BS}(x) dx, \quad (1.23)$$

$$UNC_{CRPS} = \int_{Clim_{min}}^{Clim_{max}} UNC_{BS}(x) dx. \quad (1.24)$$

Il est important de noter que les équations 1.22, 1.23 et 1.24 autorisent une représentation graphique du *CRPS* d’un système de prévision. Pour chaque niveau de GHI, on peut tracer les composants fiabilité, résolution et incertitude du *Brier score*. Les aires ainsi constituées représentent respectivement la fiabilité, la résolution et l’incertitude du *CRPS*. La partie 1.5.4 utilise cette représentation et montre comment interpréter un tel graphique.

Il convient de noter que Hersbach (2000) propose une décomposition alternative du *CRPS*. Pour cette variante, si le calcul de l’incertitude est inchangé, la fiabilité et la résolution diffèrent d’une constante D par rapport à la décomposition classique présentée ci-dessus. Il en découle que cette décomposition alternative ne conserve pas les propriétés intéressantes montrées par les équations 1.22 et 1.23 permettant de relier simplement le *Brier score* et le *CRPS*. Par conséquent, il a été choisi dans cette étude de travailler uniquement avec la décomposition classique.

Les scores linéaires

Pour la prévision de variables binaires, la définition du score négativement orienté (appelé *score linéaire naïf*)

$$\frac{1}{N} \sum_{n=1}^N |\hat{F}_{1,n} - O_n|, \quad (1.25)$$

semble naturelle. Malheureusement, il se trouve que ce score est inutilisable, car non propre. La démonstration se trouve en Annexe 1.A. Certains auteurs ont alors cherché s'il était possible de trouver des scores propres dépendant linéairement de la prévision. Bröcker et Smith (2007a) ont nommé *score propre linéaire* ou "*SPL*" un score strictement propre répondant à cette contrainte pour les prévisions de variables continues et dont la définition est :

$$SPL = \frac{1}{N} \left[\sum_{n=1}^N \int_{Clim_{min}}^{Clim_{max}} \hat{f}_n(x)^2 dx - 2\hat{f}_n(x = O_n) \right]. \quad (1.26)$$

Il est possible d'étendre la définition de ce score aux prévisions de variables multi-catégories

$$SPL = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^I \hat{F}_{i,n}^2 - 2\hat{F}_{O,n}, \quad (1.27)$$

et binaires :

$$SPL = \frac{1}{N} \sum_{n=1}^N \hat{F}_{0,n}^2 + \hat{F}_{1,n}^2 - 2\hat{F}_{O,n}. \quad (1.28)$$

Le *score propre linéaire* est cependant inutilisé en pratique pour les prévisions de variables binaires. En effet, il est dans ce cas équivalent au *Brier score*, comme démontré dans l'Annexe 1.C.

Les scores pseudo-sphériques

Les scores pseudo-sphériques sont une famille de scores propres, qui peuvent être utilisés pour des prévisions de variables binaires, multi-catégories ou continues. Ils ont été étudiés théoriquement, par exemple par Giummolè et Mameli (2018) ou Bickel (2007), mais sont néanmoins encore peu utilisés sur des cas d'applications pratiques. Gneiting et Raftery (2007) démontrent qu'ils sont strictement propres et les introduisent dans le domaine de la vérification probabiliste. Ils dépendent d'un entier naturel $\alpha \geq 2$ et sont définis pour des variables binaires par :

$$SPS_\alpha = \frac{1}{N} \sum_{n=1}^N \left[- \frac{\hat{F}_{O,n}^{\alpha-1}}{(\hat{F}_{0,n}^\alpha + \hat{F}_{1,n}^\alpha)^{\frac{\alpha-1}{\alpha}}} \right], \quad (1.29)$$

pour des variables multi-catégories par :

$$SPS_\alpha = \frac{1}{N} \sum_{n=1}^N \left[- \frac{\hat{F}_{O,n}^{\alpha-1}}{(\sum_{i=1}^I \hat{F}_{i,n}^\alpha)^{\frac{\alpha-1}{\alpha}}} \right], \quad (1.30)$$

ou pour des variables continues par :

$$SPS_\alpha = \frac{1}{N} \sum_{n=1}^N \left[- \frac{\hat{f}_n(x = O_n)^{\alpha-1}}{\left(\int_{Clim_{min}}^{Clim_{max}} |\hat{f}(x)|^\alpha dx \right)^{\frac{\alpha-1}{\alpha}}} \right]. \quad (1.31)$$

Parmi ces scores, le cas particulier $\alpha = 2$ mérite une attention particulière. Il définit le score appelé *score sphérique* (ou “*SS*”). Pour les prévisions binaires, celui-ci s’écrit :

$$SS = \frac{1}{N} \sum_{n=1}^N \left[- \frac{\hat{F}_{O,n}}{\sqrt{(\hat{F}_{0,n}^2 + \hat{F}_{1,n}^2)}} \right]. \quad (1.32)$$

Les scores logarithmiques

L’*Ignorance score* ou “*Ign*” est un autre score strictement propre provenant à l’origine de la théorie de l’information (Shannon, 1948). Il a été proposé dans le cadre de l’évaluation de prévisions probabilistes météorologiques par Roulston et Smith (2002) et est calculé de la manière suivante :

$$Ign = \frac{1}{N} \sum_{n=1}^N -\log(\hat{F}_{O,n}). \quad (1.33)$$

Ce score peut être défini de la même façon dans le cas de prévisions multi-catégories, avec un nombre d’indice plus important que 2. Il vient de sa définition que l’*ignorance score* est un score local. Benedetti (2010) parvient à la conclusion que l’*ignorance score* est le seul score propre et local pour des prévisions multi-catégories. Dans le cas de la prévision d’une variable continue, \hat{f} étant la fonction de densité de probabilité de la prévision et O l’observation correspondante, sa définition devient :

$$Ign = \frac{1}{N} \sum_{n=1}^N -\log(\hat{f}_n(x = O_n)). \quad (1.34)$$

Bernardo (1979) montre que tous les scores propres et locaux applicables aux prévisions de variables continues sont équivalents à l’*ignorance score*, et Bröcker (2009a) remarque que ce score est la limite des scores pseudosphériques lorsque $\alpha \rightarrow 1$.

L’introduction de l’*ignorance score* dans le cadre de l’évaluation des prévisions probabilistes a ouvert de nouvelles perspectives. Weijs et al. (2010) proposent deux scores pour les prévisions multi-catégories et continues qui se déduisent de l’*ignorance score* de la même façon que le *RPS* et le *CRPS* se déduisent du *Brier score*. Todter et Ahrens (2012) appellent ces scores *RIGN* (“*Ranked Ignorance Score*”) pour les prévisions multi-catégories et *CRIGN* (“*Continuous Ranked Ignorance Score*”) pour les prévisions continues. Le *RIGN* est ainsi défini par

$$RIGN = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^I Ign_{i,n}(\hat{F}, O_n), \quad (1.35)$$

où dans ce cas, la prévision associée à chaque catégorie est considérée telle une prévision binaire (O vaut 1 si l’observation appartient à la catégorie, ou 0 sinon). Pour des variables continues, le *CRIGN* se généralise à partir du *RIGN* de la même façon que le *CRPS* à partir du *RPS* (Todter, 2011) :

$$CRIGN = \frac{1}{N} \sum_{n=1}^N \int_{Clim_{min}}^{Clim_{max}} Ign_n(x) dx \quad (1.36)$$

Il est important de noter que contrairement à l'*ignorance score*, le *RIGN* et le *CRIGN* sont des scores logarithmiques globaux.

Autres scores

Le “*zero-one score*” est défini pour des prévisions binaires ou multi-catégories. Pour une prévision et une observation correspondante, la mesure de son erreur dépend uniquement des modes de la prévision \hat{F}_i (c'est-à-dire des catégories j pour lesquelles $\hat{F}_j = \max_i(\hat{F}_i)$).

Soit Nm le nombre de modes de \hat{F} , on définit :

$$ZOS_n = \begin{cases} \frac{1}{Nm} & \text{si } O \text{ appartient à un mode de } \hat{F}, \\ 0 & \text{sinon.} \end{cases} \quad (1.37)$$

Le *zero-one score* est ensuite la moyenne des mesures d'erreur, de sorte que :

$$ZOS = \frac{1}{N} \sum_{n=1}^N ZOS_n. \quad (1.38)$$

Ce score est le seul score positivement orienté étudié dans ce chapitre. Il est propre, mais pas strictement propre, et a été critiqué pour cette raison (Winkler et al., 1996). En effet, il vient immédiatement de sa définition que toutes les prévisions ayant le(s) même(s) mode(s), sans pour autant être égales, ont le même score.

Winkler (1994) et Winkler et al. (1996) proposent un nouveau score nommé *Winkler score*. Il fait partie des scores asymétriques pour les prévisions binaires, qui sont des scores pour lesquels la fonction entropie n'est pas nécessairement maximale en 0.5. Le *Winkler score* est en réalité une famille de score, car il dépend de la climatologie \bar{O} . On définit la mesure d'erreur associée au *Winkler score* par :

$$WS_n = \begin{cases} \frac{(1 - \bar{O})^2 - (1 - \hat{F}_1)^2}{\bar{O}^2 \mathbb{1}_{\hat{F}_1 < \bar{O}} + (1 - \bar{O})^2 \mathbb{1}_{\hat{F}_1 \geq \bar{O}}} & \text{si } O = 1, \\ \frac{\bar{O}^2 - \hat{F}_1^2}{\bar{O}^2 \mathbb{1}_{\hat{F}_1 < \bar{O}} + (1 - \bar{O})^2 \mathbb{1}_{\hat{F}_1 \geq \bar{O}}} & \text{si } O = 0, \end{cases} \quad (1.39)$$

où $\mathbb{1}$ est la fonction indicatrice. Dans le cas particulier où $\bar{O} = 0.5$, le *Winkler score* est équivalent au *Brier score*. Nous démontrons cette propriété dans l'Annexe 1.C.

Le *quantile score* est utilisé pour évaluer un type de prévision particulier : la prévision d'un quantile associé à un niveau de probabilité $\tau \in [0, 1]$ défini à l'avance. Autrement dit, la prévision \hat{F}_τ du niveau de probabilité τ est une estimation de $\mathcal{Q}^{-1}(\tau)$. Bentzien et Friederichs (2014) utilisent la fonction asymétrique dite de “perte Pinball”³ ρ_τ pour calculer l'erreur associée à un couple prévision/observation :

$$QS_{n,\tau} = \rho_{\tau,n}(\hat{F}_\tau, O) = \begin{cases} (O_n - \hat{F}_{\tau,n}) \times \tau & \text{si } O_n - \hat{F}_{\tau,n} \geq 0, \\ (O_n - \hat{F}_{\tau,n}) \times (\tau - 1) & \text{si } O_n - \hat{F}_{\tau,n} < 0. \end{cases} \quad (1.40)$$

³également appelée “check function”

Le *quantile score* de la prévision \hat{F}_τ associée au niveau de probabilité τ est ensuite calculé selon :

$$QS_\tau = \frac{1}{N} \sum_{n=1}^N \rho_{\tau,n}. \quad (1.41)$$

Cette définition oriente négativement le *quantile score*. Ce score est strictement propre. La démonstration en est apportée par Bentzien et Friederichs (2014). Tirant profit de la définition conditionnelle de QS , Ben Bouallègue et al. (2015) montrent qu'il est possible de décomposer le *quantile score* en fonction du signe de $O_n - \hat{F}_{\tau,n}$. Cette formulation alternative qui permet de distinguer les parts du *quantile score* dues respectivement aux sous-estimations et aux surestimations de la prévision sera utilisée dans le chapitre 4 :

$$QS_\tau = \underbrace{\frac{\tau}{N} \sum_{n=1}^N \mathbf{1}_{\hat{F}_{\tau,n} > O_n} \times (\hat{F}_{\tau,n} - O_n)}_{\text{Surestimation}} + \underbrace{\frac{\tau-1}{N} \sum_{n=1}^N \mathbf{1}_{\hat{F}_{\tau,n} < O_n} \times (\hat{F}_{\tau,n} - O_n)}_{\text{Sous-estimation}}. \quad (1.42)$$

Laio et Tamea (2007) montrent par ailleurs que le *quantile score* est lié au CRPS par la relation :

$$CRPS = 2 \times \int_{\tau=0}^{\tau=1} QS_\tau d\tau. \quad (1.43)$$

L'“*error-spread score*” (ou “*ESS*”) est un score applicable uniquement aux prévisions probabilistes des variables continues ou aux prévisions d'ensemble. Proposé par Christensen (2015), il a été développé dans le but de pouvoir être calculé sans que l'ensemble de la fonction de prévision \hat{f} soit déterminé. À la place, seules les valeurs de la moyenne de la prévision $\mu_{\hat{f}}$, de son écart-type $\sigma_{\hat{f}}$ et de son asymétrie (ou “*skewness*”) $\gamma_{\hat{f}}$ sont nécessaires. En appelant ε_n la différence entre la moyenne de la prévision $\mu_{\hat{f}_n}$ et l'observation O_n , la définition de l'*error-spread score* ESS est :

$$ESS = \frac{1}{N} \sum_{n=1}^N (\sigma_{\hat{f}_n}^2 - \varepsilon_n^2 - \varepsilon_n \sigma_{\hat{f}_n} \gamma_{\hat{f}_n})^2. \quad (1.44)$$

La démonstration du caractère propre de ce score et sa décomposition en fiabilité, résolution et incertitude se trouvent dans l'étude de Christensen (2015). En revanche, ce score n'est pas strictement propre. Par exemple, pour une prévision \hat{f} et la distribution de probabilité réelle q , si $\hat{f} \neq q$ mais $\mu_{\hat{f}} = \mu_q$, $\sigma_{\hat{f}} = \sigma_q$ et $\gamma_{\hat{f}} = \gamma_q$, on obtient l'égalité $\mathbb{E}[ESS_n(\hat{f}, q)] = \mathbb{E}[ESS_n(q, q)]$, d'où $div_{ESS}(\hat{f}, q) = 0$.

1.4.3 Carte des scores propres

Le tableau 1.1 propose une vue d'ensemble des différents scores propres qui ont été présentés dans la partie 1.4.2.

Famille	binaire	multi-catégories	continue	ensemble	quantile
Scores linéaires		<i>Score propre linéaire</i>			
Scores quadratiques	<i>Brier score</i>		<i>RPS</i>		
			<i>CRPS</i>		
Scores logarithmiques			<i>Ignorance score</i>		
			<i>RIGN</i>		
Scores pseudo-sphériques			<i>Score sphérique</i>		
			Autres scores pseudo-sphériques		
Autres scores	<i>Winkler score</i>				
	<i>Zero-one score</i> ★				
			<i>Error-spread score</i> ★		
					<i>Quantile score</i>

TABEAU 1.1 : Carte des scores propres probabilistes triés par famille et par type de prévision. Le symbole ★ indique un score qui est propre, mais non strictement propre.

1.4.4 Les scores de compétences ou “skill scores”

Pour juger des performances d’un système de prévision sur un site donné, la valeur brute d’un score n’est pas suffisante. Les variabilités de l’observation sur le site donné, les conditions de ciel, la difficulté de prévision ont des effets sur le score qui sont indépendants des performances brutes du système de prévision (Alessandrini et al., 2015). Pour quantifier la capacité d’un système de prévision à produire une prévision de qualité, il est recommandé d’utiliser un score de compétence ou “skill score”. Il permet de comparer la qualité

mesurée par n'importe quel score S d'un système de prévision avec celle d'une prévision de référence S_{ref} , en général simple à implémenter. Le *CRPSS* et le *BSS* qui sont les scores de compétence respectivement associés au *CRPS* et au *Brier score* sont souvent utilisés. En appelant $S_{parfait}$ la valeur optimale du score S , il est défini par :

$$Skill\ score = \frac{S_{ref} - S}{S_{ref} - S_{parfait}}. \quad (1.45)$$

Il est positivement orienté pour un score négativement orienté et varie entre $-\infty$ et 1. Il est défini relativement à une méthode de référence. Celle-ci doit être choisie avec soin car sa qualité définit le skill égal à 0. Le choix de ce modèle est l'objet du chapitre 2. Un skill score positif signifie que le modèle évalué est de meilleure qualité que le modèle de référence. A contrario, il est de moins bonne qualité pour un skill score négatif.

1.4.5 Outils graphiques

Alors que les scores permettent une évaluation numérique de la qualité d'un système de prévision, les outils graphiques permettent plutôt la représentation d'un attribut probabiliste préalablement choisi (calibration, finesse, fiabilité, etc.). Dans cette partie, on présente les outils couramment utilisés pour l'évaluation des prévisions de variables continues.

Le diagramme de rang et le "PIT histogram"

Le diagramme de rang, aussi appelé "rank histogram" ou "diagramme de Talagrand" est un outil graphique utilisé pour visualiser la fiabilité d'un système de prévision probabiliste. Il est de préférence utilisé pour les prévisions ensemblistes, et a été le sujet de nombreuses études théoriques par Hamill (2001) ou Candille et Talagrand (2005).

Étant donné M membres d'une prévision d'ensemble et une observation O , le rang de l'observation est défini comme la place de l'observation dans cet ensemble ordonné. Pour un nombre important de couples prévision/observation, un diagramme de rang est la représentation graphique sous forme d'histogramme du rang de l'observation. Ainsi, la $k^{ième}$ barre représente la fréquence avec laquelle l'observation occupe le $k^{ième}$ rang. Autrement dit, ce graphique met en regard de chaque intervalle d'un ensemble la proportion correspondante des observations. Chaque intervalle d'une prévision d'ensemble parfaitement fiable représente une probabilité égale d'y trouver une observation. On peut donc s'attendre à ce que le diagramme de rang d'une telle prévision constituée de M membres voit chaque intervalle être peuplé de $\frac{N}{M+1}$ valeurs observées, où N est le nombre de couples prévision/observation. C'est pourquoi la bonne fiabilité d'une prévision d'ensemble est indiquée par un diagramme de rang uniforme (où toutes les barres ont la même hauteur). En définissant S_k le nombre d'observations peuplant l'intervalle k , on peut donc calculer la déviation D de l'histogramme par rapport à un histogramme parfaitement uniforme (Candille et Talagrand, 2005 ; Talagrand et al., 1997) :

$$D = \sum_k \left(\frac{S_k - N}{M + 1} \right). \quad (1.46)$$

Un biais de la prévision peut être détecté grâce au diagramme de rang. Par exemple, des premiers rangs surpeuplés révèlent un biais positif de la médiane de l'ensemble tandis

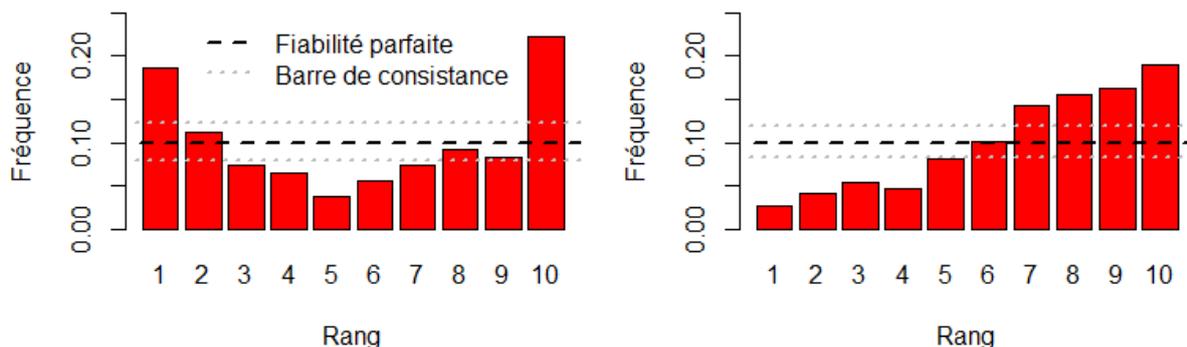
qu’une forme concave indique généralement une sur-dispersivité du modèle de prévision (les rangs centraux sont surpeuplés par les valeurs observées). Les principaux biais identifiables sont présentés sur la figure 1.2.

Cependant, deux phénomènes identifiés dans la littérature peuvent amener le diagramme de rang d’une prévision fiable à se présenter de manière non uniforme et donc mener à une mauvaise interprétation. Il s’agit du faible nombre d’observations et de l’autocorrélation temporelle :

- Pour un nombre infini de couples prévision/observation, on doit s’attendre à un diagramme de rang uniforme. En revanche, si le nombre d’observations est relativement faible, même une prévision parfaitement fiable ne pourra pas générer un diagramme uniforme. Cela est par exemple évident si le nombre d’observations est plus faible que le nombre de rangs, car tous les rangs ne peuvent alors pas être peuplés (même si la prévision est fiable). Une augmentation de la taille de l’échantillon amène le diagramme de rang d’un système de prévision fiable à s’approcher de la forme attendue.
- Pinson et al. (2010) soulignent que l’autocorrélation temporelle d’une série d’observations peut aussi entraîner la distorsion du diagramme de rang d’une prévision fiable. Pour vérifier si celle-ci joue un rôle dans la déviation constatée d’un diagramme de rang, on peut étudier les autocorrélations temporelles des observations.

À la suite de ces constatations, Bröcker et Smith (2007b) proposent d’ajouter des “barres de consistance” (ou “bandes de consistance”) sur les diagrammes de rang pour aider à leur interprétation et prendre en comptes ces effets d’échantillonnage. La barre de consistance à X % est une barre verticale représentant l’étendue des fréquences relatives de X % de prévisions fiables par construction ayant le même nombre de couples prévision/observation. Comme indiqué par Pinson et al. (2010), obtenir un diagramme pour lequel X % des fréquences mesurées se trouvent à l’intérieur de ces barres de consistance implique qu’on ne puisse pas rejeter l’hypothèse que la prévision soit fiable. Pinson et al. (2010) proposent une méthode, non détaillée ici, modifiant la procédure de construction des barres de consistance dans le but de prendre en compte l’autocorrélation temporelle des données d’observation.

Par ailleurs, Hamill (2001) alerte sur les possibilités de mauvaise interprétation, en démontrant qu’un diagramme de rang uniforme peut être généré à partir d’une prévision biaisée. Ainsi, l’uniformité ne prouve pas une bonne fiabilité.



(a) Système sous-dispersif.

(b) Système avec biais négatif.

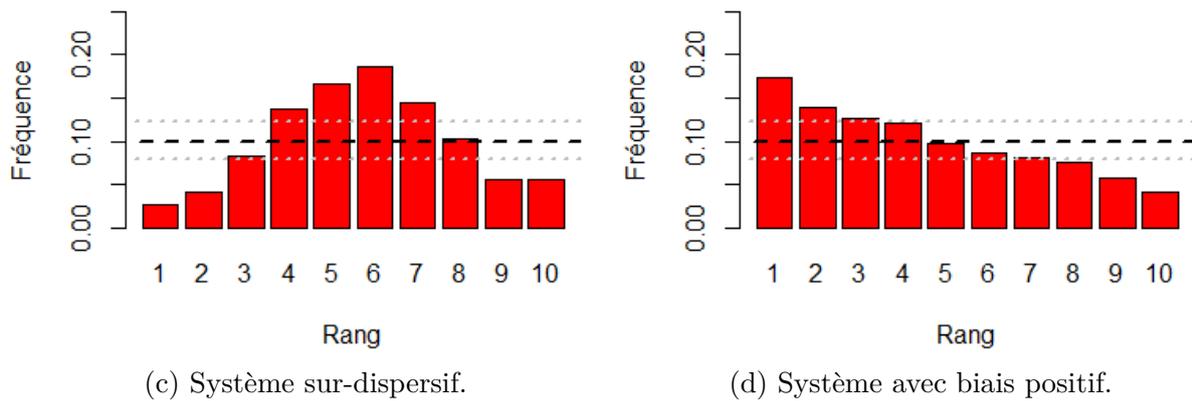


FIGURE 1.2 : Illustration des principaux biais d’une prévision d’ensemble de 9 membres identifiables à l’aide d’un diagramme de rang avec barres de consistance.

A partir de la notion de “barres de consistance”, Bröcker et Smith (2007b) proposent un nouveau graphique mettant en évidence le caractère fiable ou non de la prévision. L’idée est de tracer sur l’axe des ordonnées non plus la fréquence relative observée pour chaque rang, mais la probabilité que la fréquence observée soit plus proche de la diagonale avec un modèle de prévision fiable ayant le même nombre de couples prévision/observation. Bröcker et Smith (2007b) proposent le nom de “rank histogram on probability paper” pour un tel graphique.

Une adaptation directe du diagramme de rang pour les fonctions de prévision continues est appelée “PIT Histogram” (Dawid, 1984). C’est le même graphique, sauf que les intervalles entre membres sont remplacés par des intervalles de niveaux de probabilité. Ces intervalles doivent alors tous représenter une même masse de probabilité.

Le diagramme de fiabilité

Le diagramme de fiabilité (ou “reliability diagram”) est un autre outil utilisé pour étudier la fiabilité d’un système de prévision. Il est le pendant du diagramme de rang pour les prévisions sous forme de fonction de probabilité ou de quantiles. C’est un graphique qui montre la proportion des observations (en ordonnée) en fonction du niveau de probabilité prédit, ou probabilité nominale (en abscisse). Ce type de diagramme est introduit par Hartmann et al. (2002) puis étudié par Bröcker et Smith (2007b) ou Pinson et al. (2010).

Le défaut de fiabilité de la prévision est indiqué par la différence entre la courbe et la diagonale allant de l’origine au point de coordonnées (1, 1). Une courbe concave ou passant au-dessus de la diagonale indique que la fréquence observée est supérieure à la prévision. C’est une indication de sous-évaluation de la probabilité. Beaucoup de propriétés du diagramme de rang peuvent être transposées au diagramme de fiabilité. Par exemple, les barres de consistance peuvent être tracées aussi sur ce diagramme. La figure 1.3 montre comment le diagramme de fiabilité met en évidence les biais de prévision identifiés sur la figure 1.2.

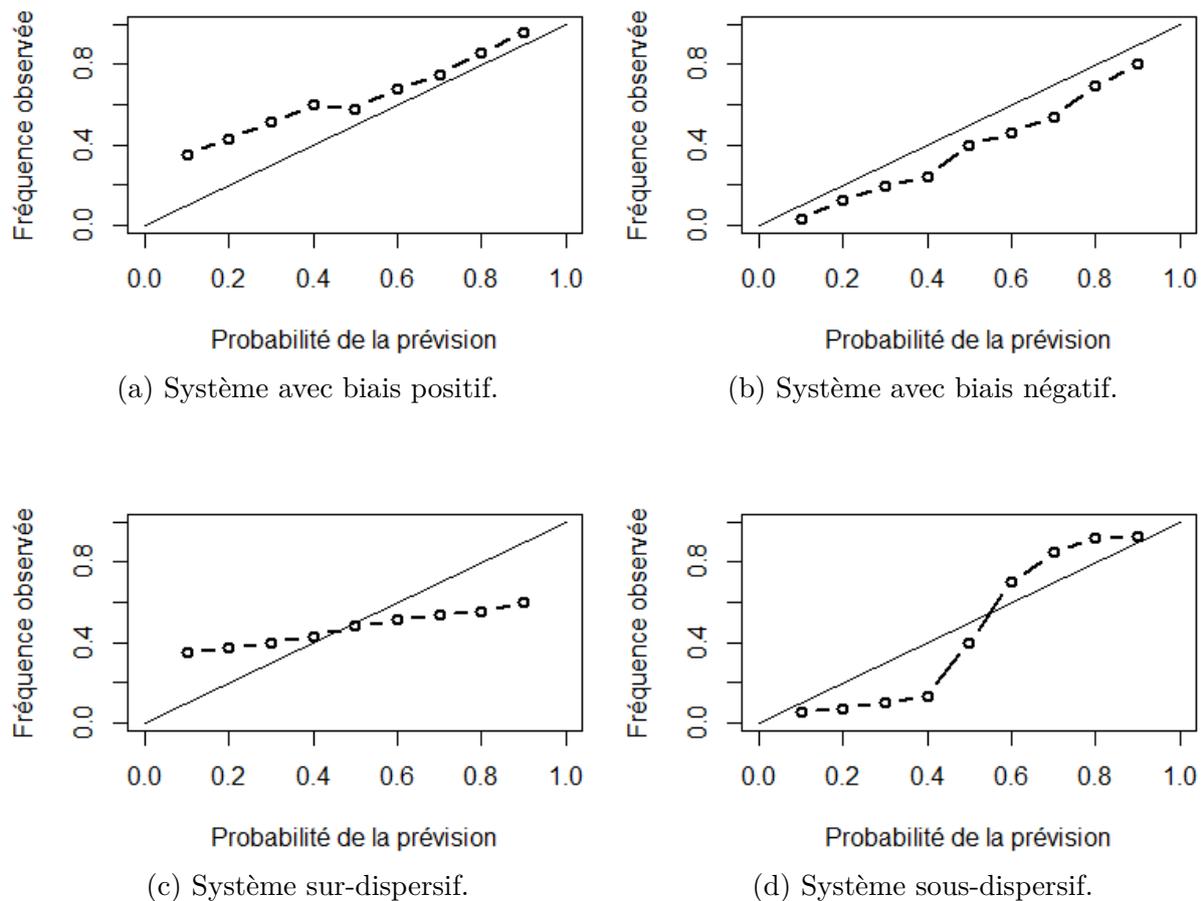


FIGURE 1.3 : Illustration des principaux biais d'une prévision identifiable à l'aide d'un diagramme de fiabilité.

Le diagramme de finesse et le diagramme PINAW

Une qualité également recherchée pour les prévisions probabilistes est la capacité à produire des intervalles de prévision étroits (ou fins). Cette propriété est appelée la "sharpness" (ou finesse) d'une prévision. Pinson et al. (2006) interprètent la fiabilité comme une propriété indispensable d'une prévision, alors que la finesse ou la résolution représentent sa valeur ajoutée. Tout graphique présentant la largeur moyenne des intervalles de prévision peut être appelé un diagramme de finesse. Normaliser la mesure de la largeur d'un intervalle par la moyenne du GHI définit le "PINAW" ("Prediction interval normalized average width"). Un diagramme PINAW est la représentation du PINAW pour l'ensemble des intervalles de probabilités centrés dont la couverture nominale varie entre 0% et 100%, c'est-à-dire l'ensemble des intervalles de la forme $[0.5 - \alpha, 0.5 + \alpha]$ avec α variant entre 0 et 0.5. Un exemple de diagramme PINAW est montré sur la figure 1.4.

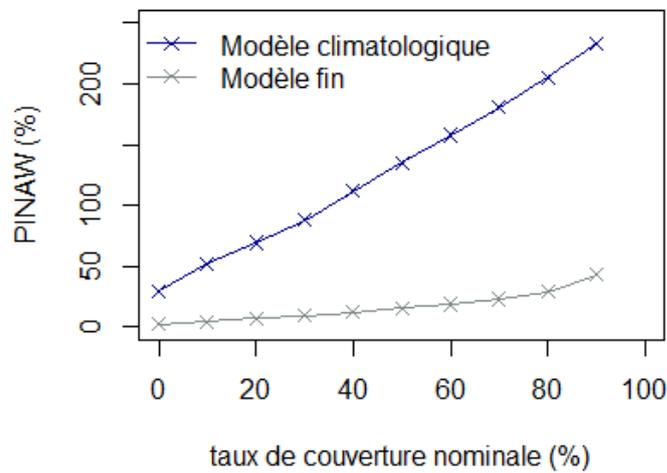


FIGURE 1.4 : Comparaison d'une prévision fine et d'une prévision de type climatologique à l'aide d'un diagramme PINAW.

Le diagramme de calibration

La calibration marginale d'une prévision est la proximité entre la fonction de répartition de prévision moyenne et la distribution moyenne des observations (cf partie 1.3). Dans le but de visualiser la correspondance entre ces deux fonctions, Gneiting et al. (2007) conseillent de tracer un diagramme de calibration, qui est la représentation sur une même figure de ces deux fonctions, ou de la différence entre ces deux fonctions. Un exemple est présenté sur la figure 1.5.

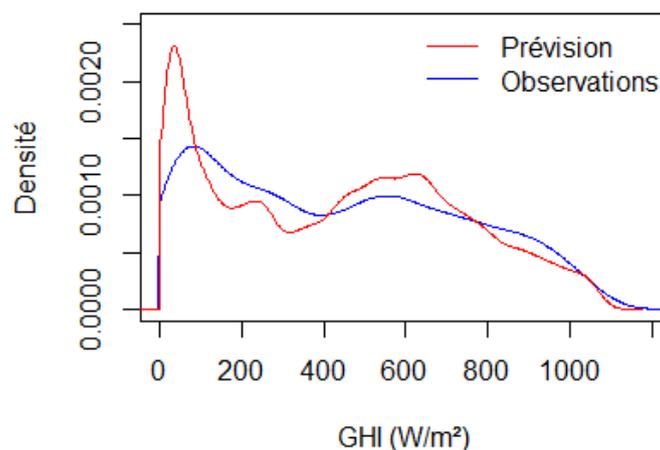


FIGURE 1.5 : Exemple d'un diagramme de calibration.

La figure 1.5 montre une prévision moyenne proche de la fonction CDF empirique des observations, excepté pour les valeurs faibles de GHI, dont la probabilité d'apparition est surévaluée.

Tableau récapitulatif

Le tableau 1.2 résume les différents outils graphiques de vérification des prévisions probabilistes vus dans cette partie. Pour chaque outil sont indiqués le type de prévision associé et l'attribut évalué. La résolution, l'incertitude et la discrimination sont exclues du tableau car aucun outil graphique dédié à ces propriétés n'a été présenté.

Outil	Type de prévision	Attribut vérifié		
		Calibration marginale	Fiabilité	Finesse
Diagramme de rang	continue ou ensemble		✓	
Diagramme de fiabilité	continue		✓	
PIT Histogram	continue		✓	
Diagramme PINAW	continue ou ensemble			✓
Diagramme de finesse	continue ou ensemble			✓
Diagramme de calibration	continue	✓		

TABLEAU 1.2 : Outils graphiques pour l'évaluation de la qualité des prévisions probabilistes de variables continues, types de prévision associés et attributs vérifiés.

1.5 Comparaison des qualités des prévisions déterministes et probabilistes

Cette section a pour but de comprendre comment les scores évaluent le gain en qualité de l'approche probabiliste par rapport à l'approche déterministe, et en particulier de quantifier le gain potentiel de l'approche probabiliste, c'est-à-dire le gain maximal qu'il est possible d'obtenir grâce à ce type de prévision. Les prévisions multi-catégories seront mises de côté dans cette partie, car elles sont peu différentes des prévisions binaires et on considère que les conclusions tirées seraient similaires. Pour comparer différents scores propres, une pratique courante dans la littérature est de tracer, étudier et comparer leur fonction entropie et leur divergence (Gneiting et Raftery, 2007). Dans ce travail, il a été décidé de procéder différemment et de traiter cette comparaison à l'aide des matrices d'espérances, un nouvel outil de représentation d'un score propre.

Tout d'abord, la partie 1.5.1 définit la matrice d'espérance d'un score et illustre sa construction sur l'exemple simple des scores de prévisions de variables binaires. Ensuite, la partie 1.5.2 étudie avec cet outil comment les scores propres de prévisions binaires identifiés dans la partie 1.4 quantifient le gain de qualité apporté par l'approche probabiliste comparativement à l'approche déterministe. Dans une troisième partie 1.5.3, la même approche sera utilisée pour quantifier les bénéfices qui peuvent être retirés des prévisions probabilistes sur les variables continues. On choisira à l'issue de cette partie de privilégier le *CRPS* pour les prévisions de variables continues et le *quantile score* pour les prévisions quantiles dans la suite de ce travail. Suite à ce choix, la partie 1.5.4 se veut être une illustration de l'utilisation pratique du *CRPS*. Elle permet en même temps de se familiariser avec les attributs des prévisions probabilistes présentés dans la partie 1.3, de mieux comprendre le fonctionnement du *CRPS*, et sa décomposition en fiabilité, incertitude et résolution.

1.5.1 Les matrices d'espérance

On définit une matrice d'espérance comme la représentation graphique, pour un score donné, de l'espérance de sa mesure d'erreur $\mathbb{E}[S_n(\hat{\mathcal{F}}, \mathcal{Q})]$ pour l'ensemble des prévisions $\hat{\mathcal{F}} \in \mathcal{E}$ et l'ensemble des distributions $\mathcal{Q} \in \mathcal{E}$. Dans cette partie, nous expliquons comment construire une telle matrice et nous illustrons ses interprétations et utilisations possibles, sur l'exemple simple des scores de prévisions de variables binaires. Pour un tel score S donné, la matrice d'espérance est construite comme suit :

1. L'espérance mathématique $\mathbb{E}[S_n(\hat{F}_1, Q_1)]$ pour chaque couple $(\hat{F}_1, Q_1) \in [0; 1]^2$ est calculée (ici avec un pas de 0.01),
2. Toutes les espérances obtenues sont représentées à l'aide d'une échelle de couleur, sur une figure où l'axe des abscisses représente la probabilité réelle Q_1 et l'axe des ordonnées la prévision \hat{F}_1 .

Il n'est donc pas nécessaire de disposer de données de prévision pour tracer ce graphique, il s'agit exclusivement un outil de représentation d'un score et des pénalisations qu'il attribue aux prévisions imparfaites. Le tableau 1.3 explicite plusieurs éléments permettant l'interprétation de ce graphique.

Élément d'interprétation	Lecture
Score obtenu à l'issue d'une observation ($O = 0$ ou $O = 1$).	Lisible sur les segments verticaux $Q_1 = 0$ et $Q_1 = 1$.
Fonction entropie \mathcal{H} de S , définie par $\mathbb{E}_{\hat{F}_1 \in [0,1]}[S_n(\hat{F}_1, \hat{F}_1)]$.	La fonction entropie associée à S est la diagonale $\hat{F}_1 = Q_1$, mise en évidence par un segment pointillé.
Caractère score propre : S est propre si $(\mathbb{E}[S_n(\hat{F}_1, Q_1)]) \geq \mathbb{E}[S_n(Q_1, Q_1)]$, pour tout $\hat{F}_1 \in [0; 1]$.	Un score est propre si pour tous les niveaux de probabilité Q_1 , l'espérance la plus faible (pour un score négativement orienté) est atteinte pour $\hat{F}_1 = Q_1$, c'est-à-dire pour le point situé sur le segment entropie. En revanche, si pour au moins un niveau de probabilité, l'espérance la plus faible n'est pas sur le segment entropie, on peut conclure que le score n'est pas propre. Un exemple de matrice de score non propre est montré sur la figure 1.6.
Décomposition du score S : $\mathbb{E}[S_n(\hat{F}_1, Q_1)] = UNC - RES + REL$.	Une matrice d'espérance permet de visualiser les composants fiabilité, résolution et incertitude d'une prévision \hat{F} . L'incertitude est l'espérance de la prévision climatologique $\mathbb{E}[S_n(\bar{O}, Q_1)]$. Cette espérance est repérée sur la représentation graphique par la couleur du point de coordonnées (Q_1, \bar{O}) . La résolution est définie comme la différence $div_S(\bar{O}, Q_1)$ entre l'espérance de la prévision climatologique et la valeur de la fonction d'entropie en Q_1 . Elle est matérialisée sur la représentation graphique par la différence de niveau de couleur entre les points (Q_1, \bar{O}) et (Q_1, Q_1) . La fiabilité est quant à elle quantifiée par la différence de couleur entre (Q_1, Q_1) et (Q_1, \hat{F}_1) . La figure 1.7 montre ces composants sur un exemple.

TABLEAU 1.3 : Interprétation de la matrice d'espérance d'un score S d'une prévision de variable binaire.

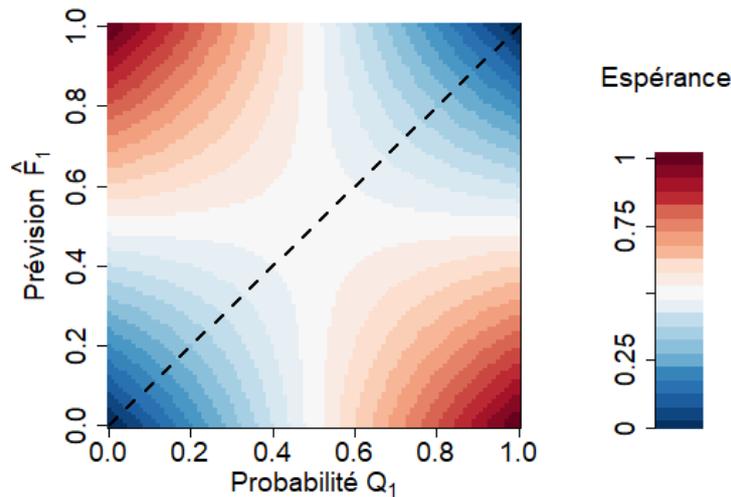


FIGURE 1.6 : Matrice d'espérance du *score linéaire naïf*.

La figure 1.6 montre la matrice d'espérance d'un score non propre : le *score linéaire naïf*, négativement orienté (cf. partie 1.4.2). On constate sur la figure 1.6 que l'espérance la plus faible pour tous les $Q_1 < 0.5$ est située en $\hat{F}_1 = 0$ et pour $Q_1 > 0.5$ en $\hat{F}_1 = 1$. La matrice d'espérance du *score linéaire naïf* montre immédiatement que celui-ci n'est pas propre, et permet de retrouver graphiquement le résultat algébrique trouvé dans l'Annexe 1.A.

La figure 1.7 est une illustration graphique de la décomposition en fiabilité, résolution et incertitude d'une prévision évaluée par le *Brier score*.

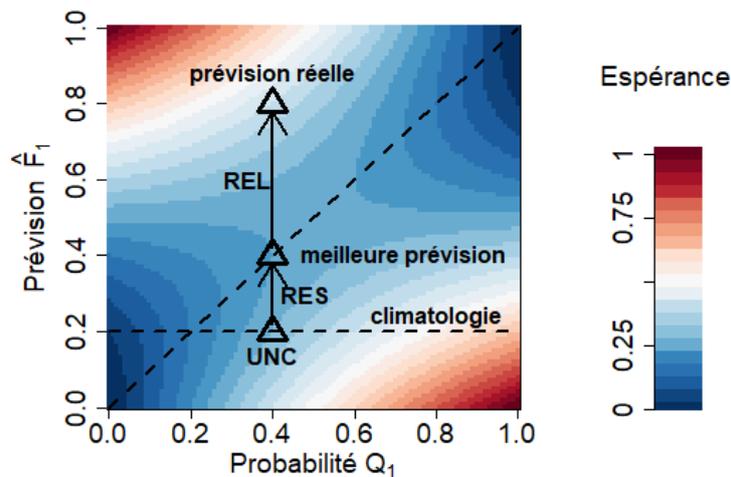


FIGURE 1.7 : Illustration de la décomposition du *Brier score* dans un cas où $\hat{F}_1 = 0.8$, $Q_1 = 0.4$ et $\bar{O} = 0.2$. *UNC* est l'espérance de la climatologie pour la probabilité réelle Q_1 . *RES* et *REL* sont les différences d'espérance (donc de niveau de couleur) entre respectivement la climatologie et la meilleure prévision et \hat{F}_1 et la meilleure prévision.

On retrouve géométriquement sur la matrice d'espérance la relation de décomposition $\mathbb{E}[BS_n(\hat{F}, Q)] = UNC - RES + REL$. Si le score est propre, la résolution est nécessairement positive car pour chaque niveau de probabilité Q_1 , l'espérance la plus faible, correspondant à la meilleure prévision, est située sur le segment entropie. Pour cette même raison, la fiabilité est nécessairement positive.

La matrice d'espérance peut aussi être étendue au diagnostic d'une prévision à l'issue d'une période de test. Cette fois, les prévisions sont regroupées en fonction de la valeur de prévision $F_{k,1}$ et l'abscisse n'est plus la probabilité Q_1 mais la fréquence observée FO_k pour le niveau de prévision $F_{k,1}$. Enfin, on place sur la matrice les points $(\hat{F}_{k,1}, FO_k)$. Une figure représentant l'ensemble de ces points existe déjà dans la littérature (Wilks, 2014) et est recommandée pour le diagnostic d'une prévision binaire. Néanmoins, la matrice d'espérance permet d'ajouter en arrière-plan le motif associé au score choisi, et de comprendre le rôle de ce choix de score dans la mesure de la qualité. En effet, le placement des points sur la matrice est uniquement dû à la prévision et ne dépend pas du score, mais ce dernier détermine complètement le motif d'arrière-plan. La combinaison du placement des points et du motif spécifique qui caractérise le score permet la compréhension de la valeur de score obtenue. Un exemple est donné sur la figure 1.8.

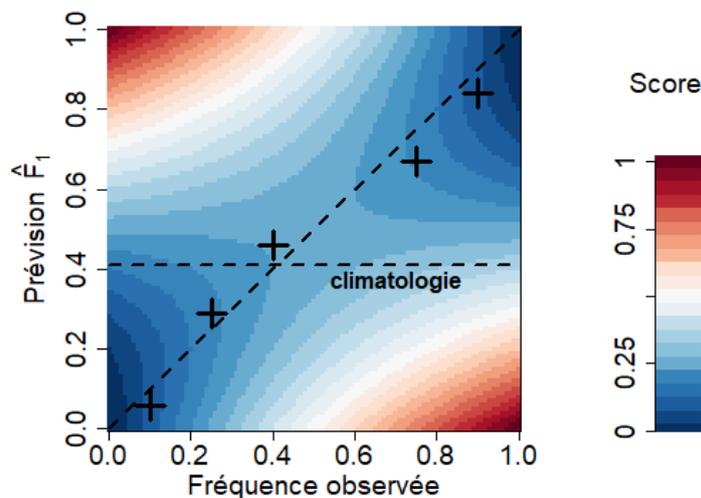


FIGURE 1.8 : Exemple d'une matrice d'espérance du *Brier score* pour une prévision réelle. Cas où le système fournit 5 niveaux de probabilité différents : 0.1, 0.25, 0.4, 0.75, 0.9 avec les fréquences observées correspondantes : 0.06, 0.29, 0.46, 0.67, 0.84.

1.5.2 Gains en qualité de l'approche probabiliste pour les scores de variables binaires

Cette partie est dédiée aux prévisions de variables binaires et mesure les gains potentiels retirés des prévisions probabilistes à l'aide des matrices d'espérance. Les scores évalués sont donc le *Brier score*, le *score sphérique*, le *zero-one score*, l'*ignorance score*, le *score propre linéaire*, le *Score linéaire naïf* et le *Winkler score*. La figure 1.9 montre les matrices d'espérance de chacun des scores étudiés dans ce travail. Le *score propre linéaire* est cependant ici mis de côté, car il est équivalent au *Brier score* (cf. Annexe 1.C), et les matrices de deux scores équivalents sont identiques.

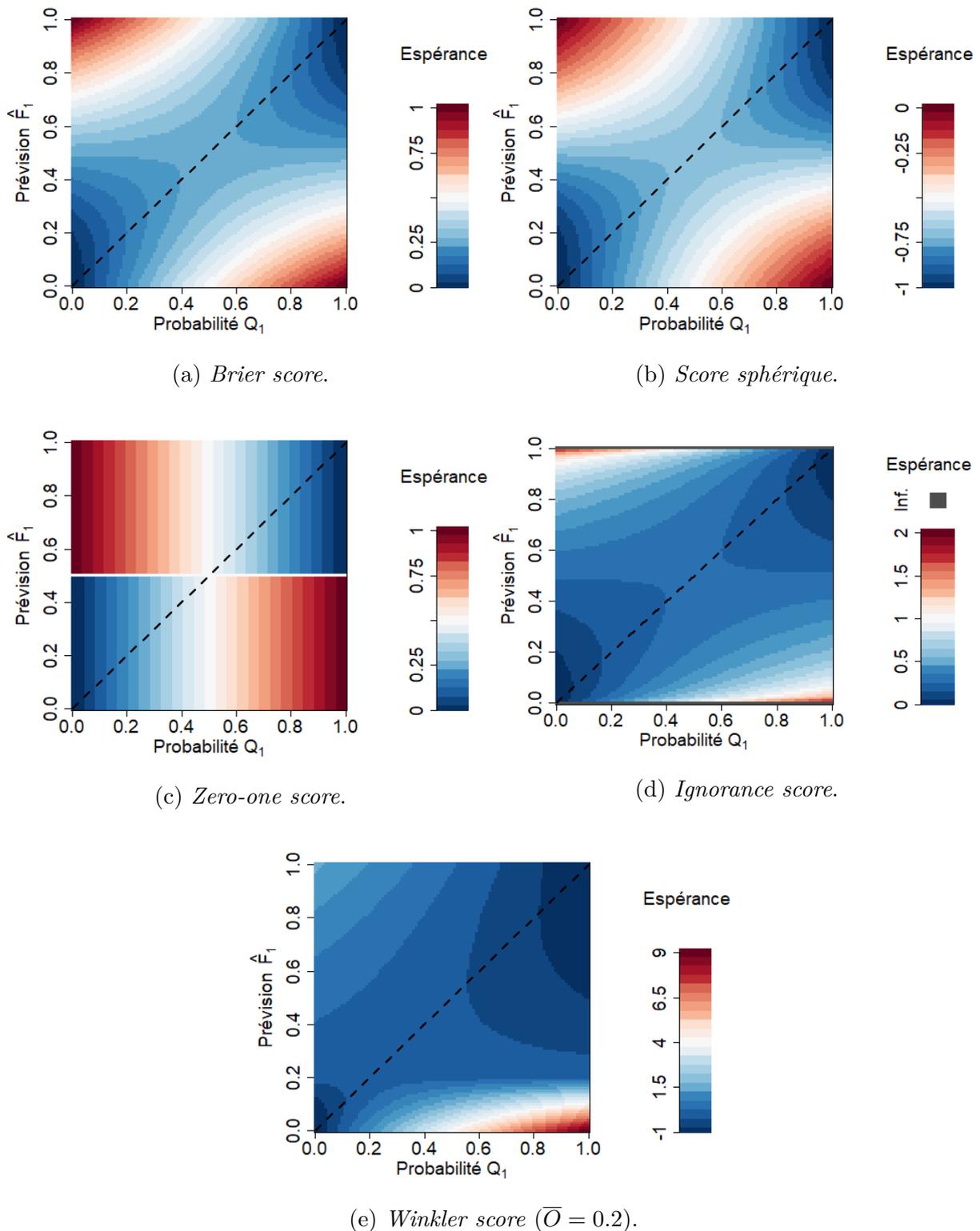


FIGURE 1.9 : Matrices d'espérance des scores applicables aux prévisions binaires.

On peut noter la forte ressemblance entre le *Brier score* et le *score sphérique*. Les trois autres scores étudiés sont en revanche très différents. En particulier, le caractère non strictement propre du *zero-one score* est mis en lumière car pour une probabilité Q_1

donnée, toutes les valeurs de \hat{F}_1 supérieures à 0.5 ont la même espérance, ainsi que toutes les valeurs de \hat{F}_1 inférieures à 0.5. L'*ignorance score* possède la propriété de pénaliser très fortement les situations pour lesquelles \hat{F}_1 et Q_1 sont très éloignées, alors qu'il y a peu de différence pour des prévisions \hat{F}_1 situées proches de 0.5. En particulier, les espérances de l'*ignorance score* sont divergentes pour $\hat{F}_1 = 0$ et $Q_1 \neq 0$ et pour $\hat{F}_1 = 1$ et $Q_1 \neq 1$. De même, le *Winkler score* pénalise fortement la prévision lorsqu'elle ne parvient pas à discriminer les cas où Q_1 est très différent de la climatologie. On peut noter que qu'il s'agit du seul score pour lequel $\mathbb{E}[S_n(\hat{F}_1 = 0.5, Q_1)]$ n'est pas constante. Au lieu de cela, c'est $\mathbb{E}[WS_n(\hat{F}_1 = \bar{O}, Q_1)]$ qui est constante, mettant en valeur le rôle clé de la climatologie pour ce score.

Ces figures permettent d'évaluer le gain de qualité de l'approche probabiliste sur l'approche déterministe. Il est déterminé par la différence entre l'espérance d'une prévision probabiliste \mathbb{E}_P et celle d'une prévision déterministe \mathbb{E}_D (pour laquelle \hat{F}_1 ne peut prendre que les valeurs 0 et 1), c'est-à-dire par la différence de couleur entre un point (\hat{F}_1, Q_1) et le point $(0, Q_1)$ ou $(1, Q_1)$. Ce gain potentiel est donc propre à chaque score et à chaque niveau de probabilité Q_1 .

L'*ignorance score* propose un gain potentiel infini, il ne permet donc pas de comparer les approches probabiliste et déterministe. De plus, on choisit d'exclure le *Winkler score* car sa définition change selon la climatologie, ce qui impacte le gain probabiliste. Pour les autres scores considérés, les gains maximums potentiels G_p de l'approche probabiliste sont calculés pour toutes les valeurs de Q_1 . Ils ont été normalisés par la différence entre l'espérance de la prévision déterministe et le score parfait du score considéré $S_{parfait}$ à la façon d'un skill score selon l'équation 1.47 puis tracés sur la figure 1.10.

$$G_p(Q_1) = \frac{\mathbb{E}_D - \mathbb{E}_P}{\mathbb{E}_D - S_{parfait}}. \quad (1.47)$$

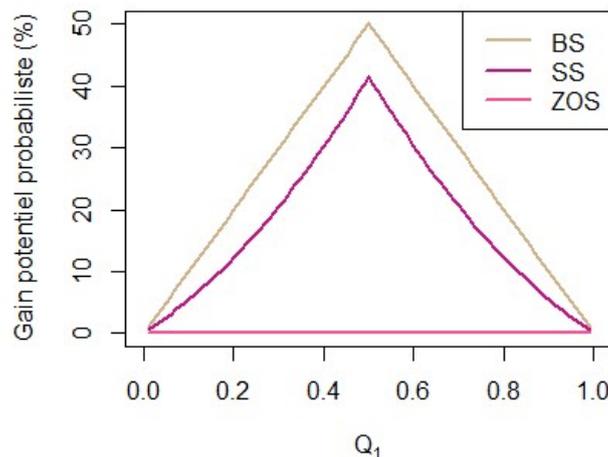


FIGURE 1.10 : Gains potentiels des scores de prévisions binaires pour toutes les valeurs de Q_1 entre 0 et 1.

Il ressort du graphique 1.10 que tous ces scores propres récompensent de manière très différente l'approche probabiliste. Le *zero-one score* est le seul score à ne pas accorder de gain à l'approche probabiliste. Cela est dû fondamentalement à son caractère non strictement propre, qui permet à la divergence d'être nulle. Les gains de qualité potentiels

des prévisions probabilistes varient selon Q_1 entre 0 et 50% pour le *Brier score* et entre 0 et 40.5% pour le *score sphérique*.

1.5.3 Gains de l'approche probabiliste pour les scores de variables continues

Dans cette partie, nous tentons d'estimer les gains maximums potentiels de l'approche probabiliste pour les scores de variables continues. Dans le cas des variables continues, les prévisions probabilistes peuvent prendre la forme d'une distribution de probabilité. On rappelle que les scores applicables aux prévisions probabilistes continues identifiés dans cette étude sont le *CRPS*, l'*ignorance score*, l'*error-spread score*, le *score propre linéaire*, le *score sphérique* et le *CRIGN*. Le *quantile score* pourrait aussi être traité dans cette partie. Néanmoins, la matrice d'espérance dépend fortement du niveau de probabilité évalué, et ce score générerait une multitude de matrices d'espérance. Elles ne sont donc pas présentées dans cette partie.

La matrice d'espérance introduite dans la partie 1.5.1 a été appliquée aux scores mentionnés ci-dessus. Le calcul du gain maximum potentiel est néanmoins plus compliqué pour les scores de variables continues, et par souci de simplicité, nous ne considérons dans cette partie que des fonctions \mathcal{Q} et $\hat{\mathcal{F}}$ gaussiennes, car une telle distribution est entièrement déterminée par sa moyenne (ici fixée par l'erreur de prévision ε), et son écart-type σ . De cette façon, seuls 4 paramètres décrivent entièrement ces deux distributions : les moyennes $\mu_{\mathcal{Q}}$ et $\mu_{\hat{\mathcal{F}}}$ et les écart-types $\sigma_{\mathcal{Q}}$ et $\sigma_{\hat{\mathcal{F}}}$. Une matrice d'espérance est alors la représentation des espérances d'un score pour l'ensemble des valeurs que peuvent prendre ces 4 paramètres. Malgré la simplification déjà effectuée, il n'est pas possible de représenter l'espérance pour l'ensemble des possibilités sur un graphique plan, et des choix doivent être effectués. Aussi, il a été décidé de fixer $\mu_{\mathcal{Q}} = 0$ et $\sigma_{\mathcal{Q}} = 0.5$. Pour chaque score, $\mathbb{E}[S(\hat{\mathcal{F}}, \mathcal{Q})]$ a été calculée pour $\mu_{\hat{\mathcal{F}}}$ et $\sigma_{\hat{\mathcal{F}}}$ variant entre 0 et 1. Avec ces hypothèses, l'approche probabiliste est caractérisée par l'utilisation de l'écart-type $\sigma_{\hat{\mathcal{F}}}$, qui est nul pour les prévisions déterministes. De cette façon, on représente sur la figure 1.11 une vue en coupe de la matrice d'espérance. On rappelle qu'un score propre est tel que son espérance est minimale pour $\mu_{\hat{\mathcal{F}}} = \mu_{\mathcal{Q}} = 0$ et $\sigma_{\hat{\mathcal{F}}} = \sigma_{\mathcal{Q}} = 0.5$. Autrement dit, le segment entropie traverse le plan de cette vue en coupe de la matrice d'espérance en $\mu_{\hat{\mathcal{F}}} = 0$ et $\sigma_{\hat{\mathcal{F}}} = 0.5$. Ce point de passage du segment entropie ($\mu_{\hat{\mathcal{F}}} = 0, \sigma_{\hat{\mathcal{F}}} = 0.5$) est mis en évidence par un point noir sur la figure 1.11. On rappelle que la différence de couleur entre une prévision probabiliste $\hat{\mathcal{F}}$ de coordonnées $(\mu_{\hat{\mathcal{F}}}, \sigma_{\hat{\mathcal{F}}})$ et une prévision déterministe $\hat{\mathcal{F}}_d$ repérée par ses coordonnées $(\mu_{\hat{\mathcal{F}}_d}, \sigma_{\hat{\mathcal{F}}_d} = 0)$ caractérise le gain potentiel de l'approche probabiliste.

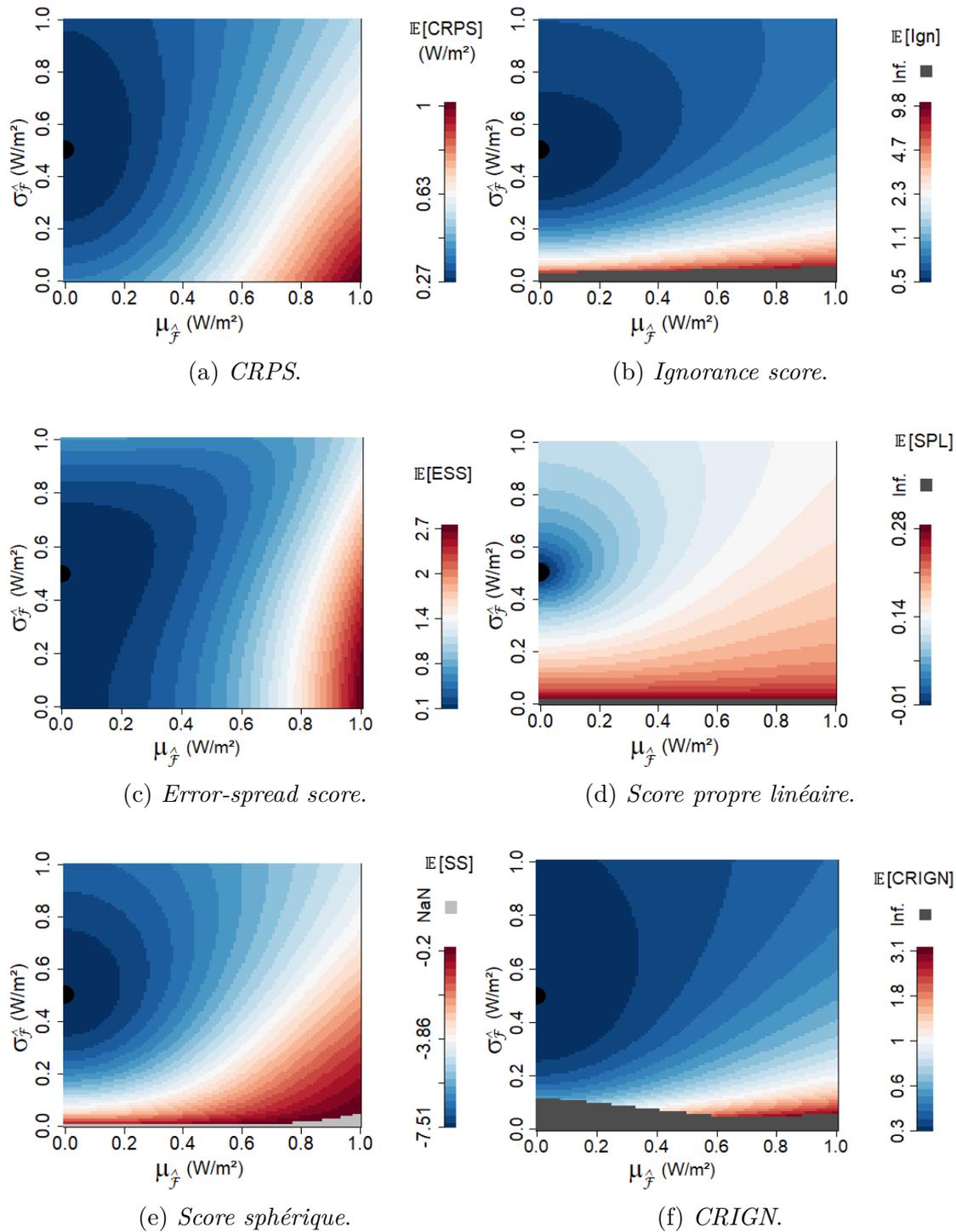


FIGURE 1.11 : Matrices d'espérance des scores de prévisions de variables continues, pour des prévisions gaussiennes dans le cas où $\mu_Q = 0$ et $\sigma_Q = 0.5$. Le passage du segment entropie dans ce plan de la matrice d'espérance est matérialisé par un point noir de coordonnées (0,0.5). Les échelles de l'*ignorance score* et du *CRIGN* sont logarithmiques. “NaN” indique une quantité non définie.

La figure 1.11 met en évidence que l'écart-type doit être utilisé par une prévision probabiliste pour retranscrire l'incertitude et tenter de maximiser la qualité. Pour tous les scores, que $\sigma_{\hat{f}}$ soit plus faible ou plus élevé que σ_Q , l'espérance la plus faible se trouve toujours pour $\mu_{\hat{f}} = \mu_Q$. En revanche, pour $\mu_{\hat{f}}$ fixée, le $\sigma_{\hat{f}}$ optimal dépend du score et de la différence $\mu_{\hat{f}} - \mu_Q$. Par exemple, pour le *score propre linéaire* et surtout l'*error-spread score*, choisir un écart-type trop important est très pénalisant et augmente l'espérance du score, même lorsque $\mu_{\hat{f}}$ est légèrement différent de μ_Q . Au contraire, les scores logarithmiques favorisent plutôt les prévisions surdispersives en pénalisant très fortement les cas où l'écart-type est trop faible. La figure 1.11 suggère donc que l'écart-type doit être choisi convenablement et en concordance avec le score utilisé pour l'évaluation de la prévision.

Les gains maximums potentiels de l'approche probabiliste par rapport à l'approche déterministe ont été calculés de la même façon que pour les scores de prévisions binaires (cf. partie 1.5.2) et sont représentés sur la figure 1.12. Pour chaque valeur de $\mu_{\hat{f}}$, la différence maximale d'espérance entre une prévision probabiliste et déterministe a été calculée. Il est clair que les valeurs d'espérance dépendent de la valeur de $|\mu_{\hat{f}} - \mu_Q|$, c'est-à-dire que le choix de $\mu_{\hat{f}}$ n'a aucun incidence, de même que le signe de $\mu_{\hat{f}} - \mu_Q$. C'est donc $|\mu_{\hat{f}} - \mu_Q|$ qui est en abscisse de la figure 1.12. On note que peu de scores permettent de comparer les approches déterministe et probabiliste car l'*ignorance score*, le *score propre linéaire*, le *score sphérique* et le *CRIGN* ont tous des espérances divergentes ou non définies pour les prévisions déterministes.

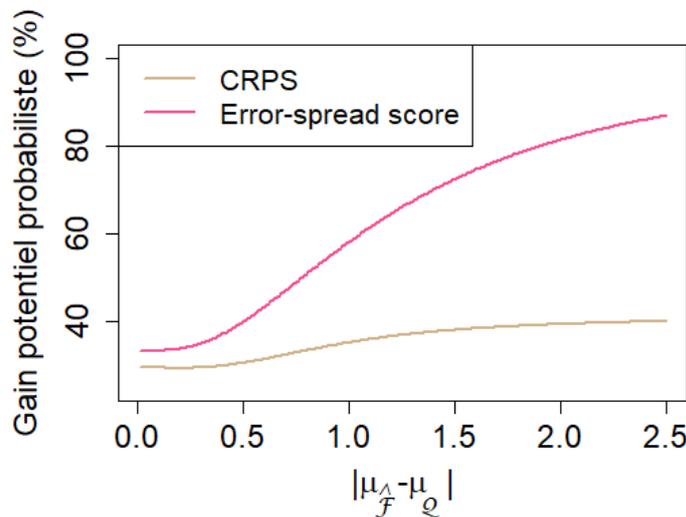


FIGURE 1.12 : Gains potentiels maximums des prévisions probabilistes gaussiennes pour les scores de variables continues avec $\sigma_Q = 0.5$ et $|\mu_{\hat{f}} - \mu_Q|$ variant entre 0 et 2.5.

La figure 1.12 prouve à nouveau l'intérêt des prévisions probabilistes dans la recherche de la meilleure qualité possible dans un contexte d'incertitude. Pour des prévisions non biaisées⁴, ces gains sont de 29,6% pour le *CRPS* et de 33,3% pour l'*ESS*. Lorsque $|\mu_{\hat{f}} - \mu_Q|$ tend vers $+\infty$, les gains potentiels tendent vers 100% pour l'*ESS* et 41% pour le *CRPS*. Ces résultats auraient pu être obtenus quelle que soit la valeur de σ_Q prise comme hypothèse, à l'exception de $\sigma_Q = 0$. Cependant, il convient de noter que ces valeurs doivent être

⁴au sens probabiliste, c'est-à-dire telles que $\mu_{\hat{f}} = \mu_Q$

considérées uniquement comme des ordres de grandeur car ces gains ont été calculés en considérant des hypothèses fortes (prévisions uniquement gaussiennes). C'est pourquoi il est plus difficile de déterminer pour ces scores le gain maximal potentiel de l'approche probabiliste que pour les scores de prévisions binaires.

En ce qui concerne le *quantile score* dont les matrices d'espérance n'ont pas été tracées, les gains maximums dépendent du niveau de probabilité τ évalué. Cependant, il est assez clair que ces gains peuvent varier de 0% pour le cas $\tau = 0.5$ (car la prévision probabiliste ne modifie pas la meilleure estimation de ce niveau de probabilité) à 100% pour les cas $\tau = 0$ et $\tau = 1$.

Pour tenter de lier les gains en qualité et les gains en valeur, on souhaite garder uniquement les scores identifiant des gains potentiels finis. Dans la partie 1.5.3, il a été établi que le *CRPS*, l'*ESS* et le *quantile score* répondent à ce critère. Prenant en compte que l'*ESS* est critiquable car il n'est pas strictement propre, nous choisissons dans la suite de cette étude de travailler préférentiellement avec le *CRPS* et le *quantile score*.

1.5.4 D'une mauvaise prévision à la prévision parfaite

Étant donné les choix faits dans la partie 1.5.3, nous présentons dans cette partie une étude qui a pour objectif d'expliquer comment peut être interprétée la décomposition graphique du *CRPS*, détaillée dans la partie 1.4.2. À partir d'une mauvaise prévision de départ déterministe et biaisée, on étudie comment évoluent les différents attributs d'une prévision probabiliste présentés dans la section 1.3 lorsque la prévision devient parfaite. 8 prévisions différentes générées artificiellement, au fil desquelles la qualité s'améliore continuellement, constituent cette étude. Gneiting et al. (2007) expliquent que la fiabilité est l'attribut le plus essentiel, qu'il faut améliorer en priorité, puis qu'il faut tenter par la suite et dans la mesure du possible de parfaire la qualité grâce à la finesse et la résolution. Pour suivre cette approche, les cas n°1 à 5 montrent dans un premier temps une amélioration de la fiabilité, et les cas n°5 à 8 présentent une amélioration de la résolution tout en gardant une fiabilité parfaite. Le tableau 1.5 et la figure 1.13 décrivent les 8 prévisions. Les observations sont quant à elles constituées de la série temporelle de GHI mesurée sur le site de St-Pierre (île de la Réunion) pour les années 2012 et 2013, dont les principales caractéristiques sont présentées dans le tableau 1.4

Moyenne (W/m^2)	424
Médiane (W/m^2)	391
UNC_{CRPS} (W/m^2)	172

TABLEAU 1.4 : Caractéristiques des mesures de GHI de Saint-Pierre, années 2012 et 2013.

Cas	Type de prévision	Caractéristiques
1	Déterministe	Prévision biaisée constante (212 W/m^2).
2	Déterministe	Prévision non biaisée constante (391 W/m^2).
3	Probabiliste	Prévision constante centrée, nettement trop fine, fiable seulement pour les valeurs de GHI médianes.
4	Probabiliste	Prévision constante centrée, légèrement trop fine, non fiable seulement pour les valeurs de GHI extrêmes.
5	Probabiliste	Prévision climatologique, fiable mais non résolue.
6	Probabiliste	Prévision fiable peu résolue, qui est parvenue à trier les observations en 2 groupes mutuellement exclusifs : M1 ($O < 424 \text{ W/m}^2$) et M2 ($O > 424 \text{ W/m}^2$).
7	Probabiliste	Prévision fiable très résolue, qui est parvenue à trier les observations en 4 groupes mutuellement exclusifs : Q1 ($O < 148 \text{ W/m}^2$), Q2 ($148 \text{ W/m}^2 < O < 391 \text{ W/m}^2$), Q3 ($391 \text{ W/m}^2 < O < 663 \text{ W/m}^2$) et Q4 ($663 \text{ W/m}^2 < O$).
8	Déterministe	Prévision parfaite.

TABLEAU 1.5 : Présentation des caractéristiques des 8 prévisions du cas d'étude.

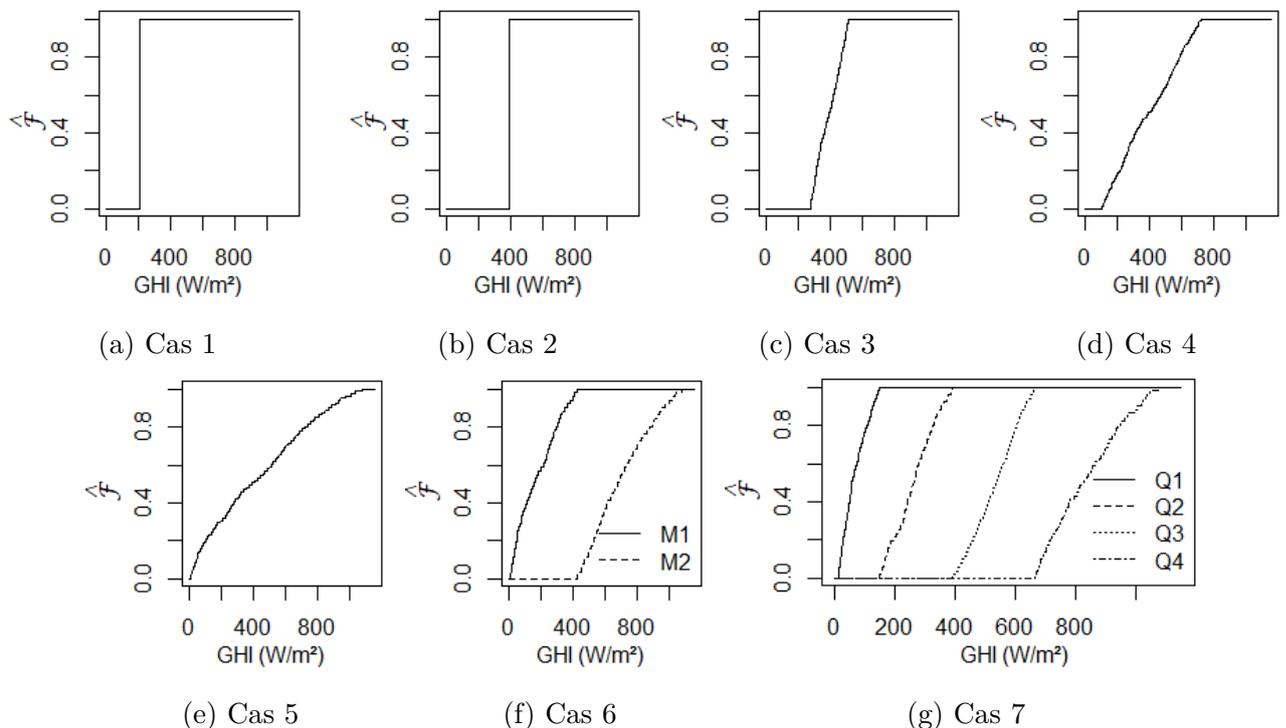


FIGURE 1.13 : Fonctions CDF des 7 premières prévisions. Celle du cas 5 est la CDF climatologique.

Les figures 1.14 et 1.15 montrent, pour chacun des 8 cas de prévision, la représentation graphique du *CRPS* introduite dans la section 1.4.2 et mettant en évidence les différents composants fiabilité, résolution et incertitude. Suivant cette décomposition, ces composants sont visualisés par des aires (ici jaune pour la fiabilité, bleue pour la résolution et hachurée pour l'incertitude) sous la courbe du *Brier score* $BS(x)$ suivant l'équation $BS(x) = REL_{BS}(x) + UNC_{BS}(x) - RES_{BS}(x)$. Les composants incertitude et fiabilité s'ajoutent alors que la résolution vient diminuer le *Brier score*. On rappelle que l'incertitude visualisée par l'aire hachurée est constante entre les différents cas.

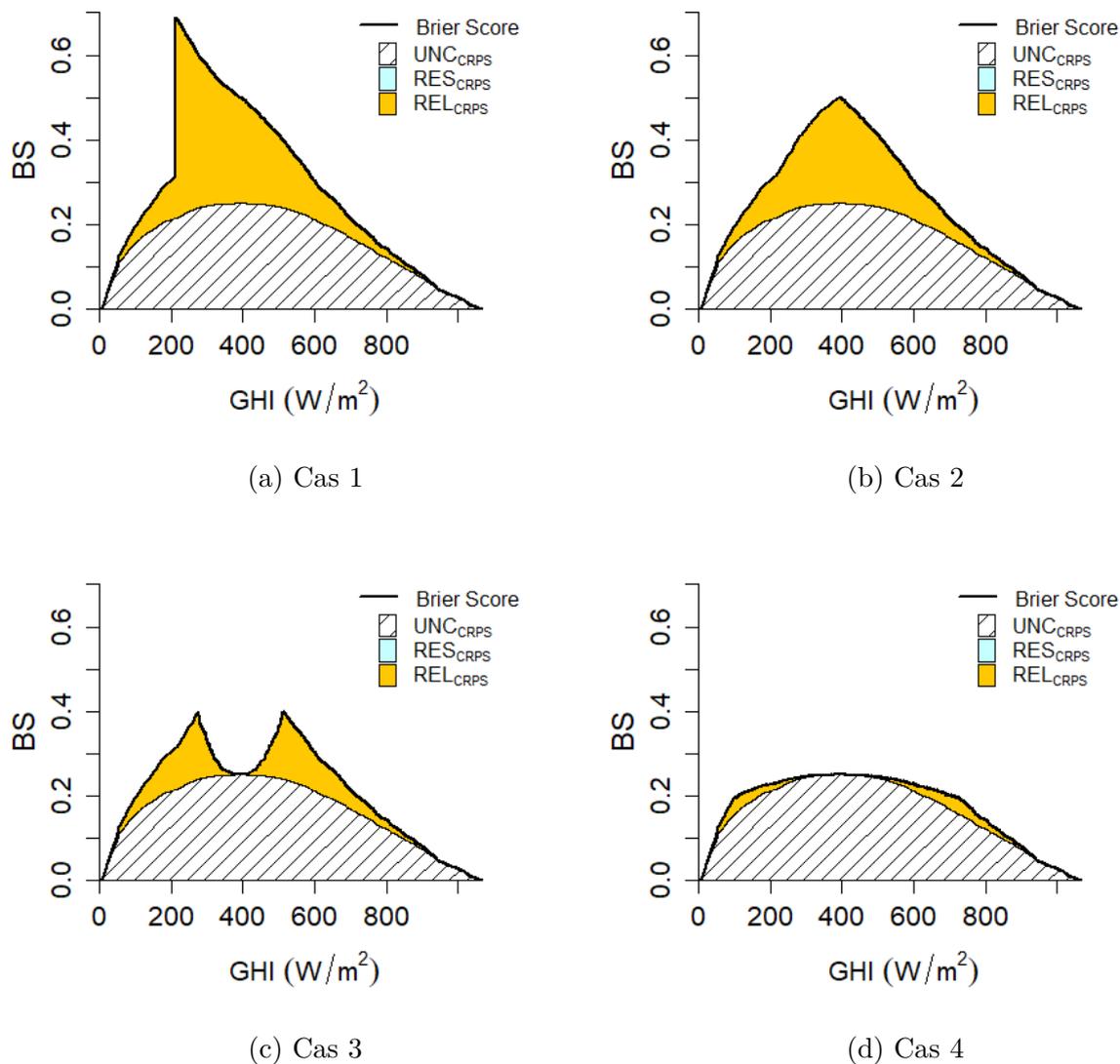


FIGURE 1.14 : Amélioration de la fiabilité.

Les résolutions de tous ces cas sont nulles car les prévisions sont constantes. Pour le cas n°1, la fiabilité est très mauvaise entre la valeur de prévision déterministe et la moyenne climatologique. La comparaison entre les cas n°1 et 2 montre que la fiabilité est très sensible au biais. Pour les cas n°3 et 4, la moyenne de la prévision reste non biaisée mais la prévision devient probabiliste, alors que la finesse diminue. La valeur de REL_{CRPS}

diminue alors aussi progressivement, en commençant par les valeurs médianes de GHI.

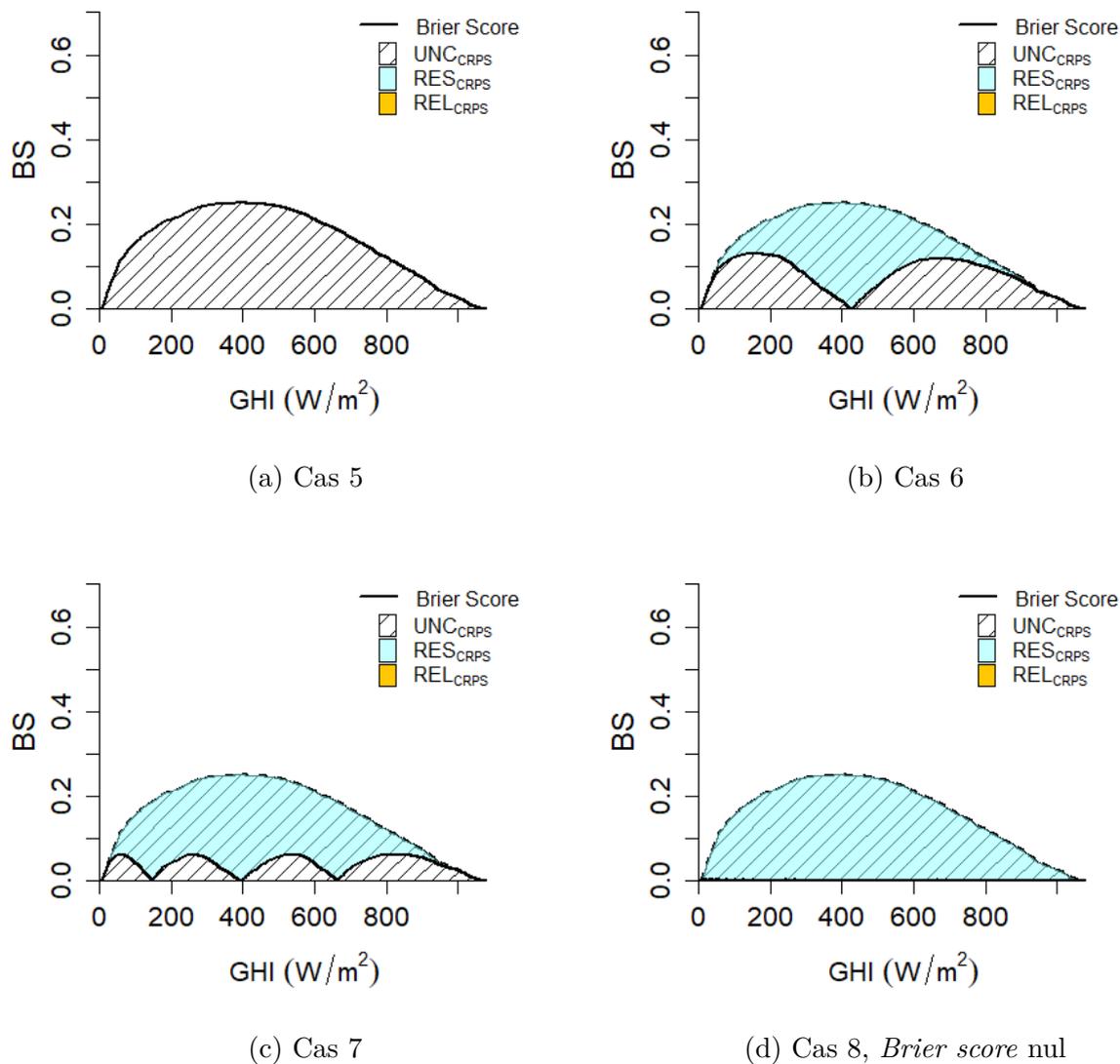


FIGURE 1.15 : Amélioration de la résolution.

Les cas n°5 à 8 sont tous parfaitement fiables ($REL_{CRPS} = 0$). La figure 1.15a montre que la prévision climatologique est fiable, mais non résolue. Le $CRPS$ est alors égal à l'incertitude UNC_{CRPS} . Les prévisions n°5 et 6 qui parviennent à produire différentes CDF selon la valeur finale de l'observation sont récompensées à travers le composant résolution RES_{CRPS} . Enfin, on voit que la prévision déterministe parfaite parvient à capter toute la variabilité de la série temporelle d'observation. Dans ce cas, on obtient une résolution égale à l'incertitude et le $CRPS$ est nul (*Brier score nul* pour toutes les valeurs de GHI).

La figure 1.16 présente l'évolution du $CRPS$ et des propriétés des prévisions probabilistes présentées dans la partie 1.3 (incertitude, fiabilité, résolution, finesse) pour les différents cas présentés dans les figures 1.14 et 1.15.

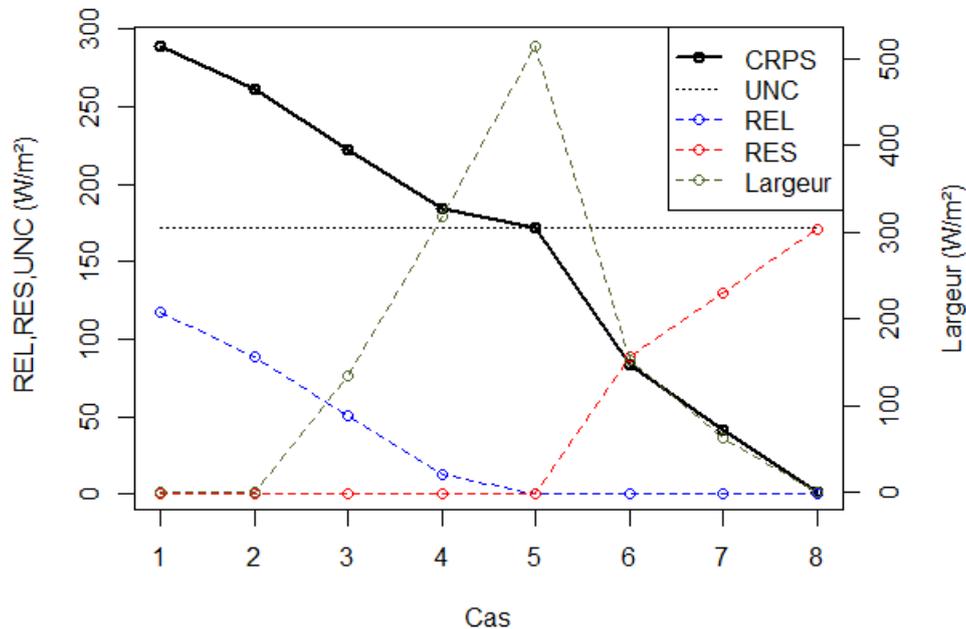


FIGURE 1.16 : *CRPS*, incertitude, fiabilité, résolution et finesse des différents systèmes de prévision présentés dans le tableau 1.5. La finesse est matérialisée par la largeur (en W/m^2) de l'intervalle de probabilité $[0.25-0.75]$.

On peut observer une décroissance progressive du *CRPS*, confirmant l'amélioration continue de la qualité au fil des différentes prévisions. Du cas n°1 au cas n°5, l'amélioration de la fiabilité explique la diminution du *CRPS*. En revanche, du cas n°5 au cas n°8, c'est l'augmentation de la résolution (de 0 jusqu'à *UNC*) qui est la raison de la diminution du *CRPS* et de l'amélioration de la qualité. La largeur de l'intervalle de prévision $[0.25-0.75]$ est nulle pour les prévisions déterministes non fiables (cas n°1 et n°2), puis augmente à mesure que la fiabilité s'améliore jusqu'à devenir maximale pour la prévision climatologique. Elle diminue ensuite quand la résolution augmente, pour redevenir nulle pour la prévision parfaite déterministe. Cette figure met clairement en évidence que la finesse n'est pas nécessairement corrélée à la résolution. En effet, elle change considérablement entre les cas n°2 et n°5 sans que la résolution ne soit modifiée.

1.6 Conclusions

L'objectif de ce chapitre est de mesurer le gain potentiel de qualité des prévisions probabilistes et d'identifier un ou plusieurs scores probabilistes quantifiant la qualité qui pourraient potentiellement permettre dans les futurs chapitres de mettre en relation les gains en qualité et en valeur. Il débute par une étude des attributs des prévisions probabilistes. À travers une revue bibliographique sur les outils de vérification des prévisions probabilistes, différents scores pouvant répondre à l'objectif ont été identifiés. Ce travail a pour conclusion l'établissement d'une carte des scores propres probabilistes et d'un tableau récapitulatif des outils graphiques de vérification qui constituent une contribution

importante de ce chapitre.

Ensuite, un nouvel outil de représentation d'un score nommé matrice d'espérance a été présenté. Il permet de visualiser graphiquement plusieurs propriétés des scores : l'équivalence, l'entropie, le caractère propre et strictement propre, la décomposition d'une prévision et le gain de qualité potentiel des prévisions probabilistes par rapport aux prévisions déterministes. Il a été utilisé pour quantifier les gains potentiels de l'approche probabiliste pour les différents scores identifiés dans la revue de la littérature. Les gains potentiels de l'approche probabiliste obtenus dans cette partie peuvent être résumés sur le tableau 1.6, mais il faut rappeler qu'il est plus complexe de conclure pour les scores de variables continues que pour les scores de variables binaires, car des hypothèses fortes ont dues être posées (seules des lois normales ont été considérées) pour établir les matrices d'espérance dont ces résultats sont issus.

		Score	Gain potentiel probabiliste
Type de variables	binaire	<i>BS</i>	0-50 %
		<i>SS</i>	0-41.5 %
		<i>ZOS</i>	0 %
		<i>Ignorance score</i>	∞
	continue	<i>CRPS</i>	29.4-40.5%
		<i>Ignorance score</i>	∞
		<i>ESS</i>	33.3-100 %
		<i>SPL</i>	∞
		<i>SS</i>	NaN
		<i>CRIGN</i>	∞
	<i>Quantile score</i>	0-100 %	

TABLEAU 1.6 : Synthèse des gains potentiels maximums de l'approche probabiliste par rapport à l'approche déterministe pour les différents scores considérés, lorsque les fonctions PDF réelle de la variable cible et de la prévision sont gaussiennes. "NaN" indique une quantité non définie.

Il a ensuite été mis en évidence que le *CRPS* et le *quantile score* seront les scores privilégiés par la suite pour étudier les gains en qualité et éventuellement les comparer avec les gains en valeur. En effet, ils accordent des gains potentiels probabilistes finis, sont strictement propres et utilisables pour des prévisions de variables continues.

À la suite de ce choix, une étude de l'évolution des différents composants du *CRPS* pour différentes prévisions a été proposée afin d'aider à la compréhension de ce score et de sa décomposition.

Annexes

1.A Démonstration du caractère propre des scores binaires de type “fonction puissance”

Dans cette partie nous étudions le caractère propre de toutes les mesures d’erreur de scores de variables binaires qui s’écrivent sous la forme :

$$S_n = |\hat{F}_1 - O|^\alpha, \quad (1.48)$$

avec $\alpha \in \mathbb{N}$. Cette démonstration s’applique en particulier au *score linéaire naïf* (cas où $\alpha = 1$) et au *Brier score* ($\alpha = 2$).

Nous allons montrer que la caractéristique “score propre” (l’espérance est minimale pour $\hat{F}_1 = Q_1$, où (Q_0, Q_1) est la probabilité réelle) apparaît uniquement pour $\alpha = 2$ (ce qui correspond à la définition du *Brier score*). Le théorème de transfert indique que pour une variable discrète X et toute fonction ϕ , on a :

$$\mathbb{E}[\phi(X)] = \phi(0)Q_0 + \phi(1)Q_1, \quad (1.49)$$

ce qui donne, appliqué à la mesure S_n et à une probabilité réelle $q = Q_1$:

$$\mathbb{E}[S_n(\hat{F}_1, Q_1)] = (1 - \hat{F}_1)^\alpha Q_1 + (1 - Q_1)\hat{F}_1^\alpha. \quad (1.50)$$

Pour trouver la valeur de \hat{F}_1 qui minimise $\mathbb{E}[S_n(\hat{F}_1, Q_1)]$, on peut dériver cette expression par rapport à \hat{F}_1 , ce qui donne :

$$\frac{d\mathbb{E}(S_n)}{d\hat{F}_1} = -\alpha Q_1(1 - \hat{F}_1)^{\alpha-1} + \alpha(1 - Q_1)\hat{F}_1^{\alpha-1} = 0. \quad (1.51)$$

Pour $\alpha \geq 1$, Cette fonction est elle-même dérivable, avec :

1. Pour $\alpha = 1$:

$$\frac{d^2\mathbb{E}[S_n]}{d\hat{F}_1^2} = 0. \quad (1.52)$$

Dans ce cas, la dérivée $\frac{d\mathbb{E}(S)}{d\hat{F}_1}$ est monotone et vaut $1 - 2Q_1$, ce qui ne dépend pas de \hat{F}_1 . On peut séparer ce cas en 3 :

- $Q_1 < 0.5$, auquel cas $\frac{d\mathbb{E}[S_n]}{d\hat{F}_1} < 0$, donc le minimum de $\mathbb{E}[S_n]$ est atteint pour $\hat{F}_1 = 0$, avec $\mathbb{E}[S(0, Q_1)] = Q_1$;
- $Q_1 > 0.5$, auquel cas $\frac{d\mathbb{E}[S_n]}{d\hat{F}_1} > 0$, donc le minimum de $\mathbb{E}[S_n]$ est atteint pour $\hat{F}_1 = 1$, avec $\mathbb{E}[S(1, Q_1)] = 1 - Q_1$;
- $Q_1 = 0.5$, auquel cas $\frac{d\mathbb{E}[S_n]}{d\hat{F}_1} = 0$, $\mathbb{E}[S_n]$ est alors constante de valeur : $\mathbb{E}[S_n] = \frac{1}{2}$.

Pour chacun de ces 3 cas, l’espérance ne dépend pas de \hat{F} , et on n’a pas la relation définissant un score propre : $N = 1$ ne génère donc pas un score propre.

2. Pour $\alpha > 1$:

$$\frac{d^2\mathbb{E}[S_n]}{d\hat{F}_1^2} = \alpha(\alpha - 1)(1 - \hat{F}_1)^{\alpha-2}Q_1 + \alpha(\alpha - 1)(1 - Q_1)\hat{F}_1^{\alpha-2} = 0. \quad (1.53)$$

Cette dérivée seconde est elle-même continue et strictement positive, ce qui justifie que le minimum de $\mathbb{E}[S_n]$ soit solution de :

$$\frac{d\mathbb{E}[S_n]}{d\hat{F}_1} = 0, \quad (1.54)$$

soit :

$$(1 - \hat{F}_1)^{\alpha-1}Q_1 - (1 - Q_1)\hat{F}_1^{\alpha-1} = Q_1[(1 - \hat{F}_1)^{\alpha-1} + \hat{F}_1^{\alpha-1}] - \hat{F}_1^{\alpha-1} = 0. \quad (1.55)$$

Donc on cherche finalement \hat{F}_1 telle que :

$$Q_1 = \frac{\hat{F}_1^{\alpha-1}}{((1 - \hat{F}_1)^{\alpha-1} + \hat{F}_1^{\alpha-1})}. \quad (1.56)$$

Un score propre étant minimum pour $\hat{F}_1 = Q_1$, la seule solution possible est de choisir $\alpha = 2$, ce qui correspond à la définition de la mesure d'erreur du *Brier score* et qui justifie son utilisation. Dans ce cas, on trouve :

$$\mathbb{E}[BS_n] = (1 - \hat{F}_1)^2Q_1 + (1 - Q_1)\hat{F}_1^2 = (1 - Q_1)Q_1 + (\hat{F}_1 - Q_1)^2. \quad (1.57)$$

De plus, l'entropie associée au *Brier score* s'écrit donc $\mathcal{H}(q) = Q_1(1 - Q_1)$ soit finalement :

$$\mathbb{E}(BS_n[\hat{F}_1, Q_1]) = \mathcal{H}(Q_1) + (\hat{F}_1 - Q_1)^2. \quad (1.58)$$

L'espérance du score peut donc être décomposée en la somme de 2 termes : l'entropie qui ne dépend que de la probabilité réelle Q_1 , et l'écart quadratique entre la probabilité réelle et la probabilité de prévision.

1.B Décomposition du *Brier score*

Le *Brier score* est défini par :

$$BS = \frac{1}{N} \sum_{n=1}^N (\hat{F}_{1,n} - O_n)^2. \quad (1.59)$$

On introduit \bar{O} , la moyenne climatologique de l'évènement considéré. On a donc :

$$\bar{O} = \frac{1}{N} \sum_{n=1}^N O_{k,n}. \quad (1.60)$$

On considère maintenant l'ensemble des K valeurs différentes de prévision données par le modèle. On peut écrire \hat{F}_1 sous la forme d'une famille de K membres :

$$(\hat{F}_1) = (\hat{F}_{1,k}), \text{ avec } 1 \leq k \leq K. \quad (1.61)$$

On introduit alors l'observation moyenne \bar{O}_k des N_k observations lorsque $F_1 = F_{1,k}$:

$$\bar{O}_k = \frac{1}{N_k} \sum_{n=1}^{N_k} O_n. \quad (1.62)$$

On définit ensuite BS_k , le score de chaque membre de la famille ($\hat{F}_{1,k}$) :

$$BS_k = \frac{1}{N_k} \sum_{n=1}^{N_k} (\hat{F}_{1,k,n} - O_{k,n})^2, \quad (1.63)$$

et, en développant BS_k :

$$BS_k = \hat{F}_{1,k}^2 - \frac{1}{N_k} \times \left(2\hat{F}_{1,k} \times \sum_{n=1}^{N_k} O_{k,n} - \sum_{n=1}^{N_k} O_{k,n}^2 \right), \quad (1.64)$$

$$BS_k = \hat{F}_{1,k}^2 - \frac{1}{N_k} (2\hat{F}_{1,k} - 1) \sum_{n=1}^{N_k} O_{k,n} = (\hat{F}_{1,k} - \bar{O}_k)^2 + \bar{O}_k(1 - \bar{O}_k), \quad (1.65)$$

en utilisant $O^2 = O$. En effet, O n'a que 2 valeurs possibles : 0 ou 1.

Or, le *Brier score* est la moyenne pondéré des BS_k :

$$BS = \frac{1}{N} \sum_{k=1}^K N_k \times BS_k. \quad (1.66)$$

On trouve donc :

$$BS = \underbrace{\frac{1}{N} \sum_{n=1}^N N_k \times (\hat{F}_{1,k} - \bar{O}_k)^2}_{REL} + \underbrace{\frac{1}{N} \sum_{n=1}^N N_k \times (\bar{O}_k \times (1 - \bar{O}_k))}_{REF}. \quad (1.67)$$

On considère alors les 2 membres de la somme ci-dessus : le membre de gauche est la fiabilité du modèle de prévision (*REL*). Il s'agit d'une mesure négativement orientée qui permet de déterminer si le modèle donne des probabilités proches des probabilités réelles. Le membre de droite est appelé "refinement" (*REF*). Il peut être décomposé comme suit (où \bar{O} est la moyenne climatologique) :

$$REF = \frac{1}{T} \sum_{k=1}^K N_k (O_k \times (1 - O_k)) = \bar{O} - \frac{1}{T} \sum_{k=1}^K (N_k \times O_k^2), \quad (1.68)$$

$$REF = \bar{O} - \bar{O}^2 - \left(\frac{1}{N} \sum_k N_k \times O_k^2 - 2\bar{O}^2 + \bar{O}^2 \right), \quad (1.69)$$

$$REF = \underbrace{(1 - \bar{O})\bar{O}}_U - \underbrace{\frac{1}{N} \sum_k N_k (O_k - \bar{O})^2}_{RES}. \quad (1.70)$$

Le "refinement" est ainsi lui-même la différence entre deux termes : l'incertitude *UNC* et la résolution *RES*. On en déduit finalement la décomposition canonique du *Brier score* :

$$BS = REL + U - RES. \quad (1.71)$$

On peut noter que dans le cas spécifique du *Brier score*, certaines décompositions alternatives ont été proposées. Elles sont souvent définies à partir de la décomposition canonique, et basées sur des sous-décomposition de la fiabilité ou de la résolution (Stephenson et al., 2008 ; Yates, 1982).

1.C Équivalences du *Brier score*, du *score propre linéaire* et du *Winkler score*

On rappelle les expressions des mesures d'erreur du *Brier score* et du *score propre linéaire* :

$$BS_n(\hat{F}_1, O) = (\hat{F}_1 - O)^2, \quad (1.72)$$

$$SPL_n(\hat{F}_1, O) = \hat{F}_0^2 + \hat{F}_1^2 - 2\hat{F}_O. \quad (1.73)$$

Si $O = 1$, on trouve directement que :

$$SPL_n(\hat{F}_1, O = 1) = (1 - \hat{F}_1)^2 + \hat{F}_1^2 - 2\hat{F}_1 = 1 - 4\hat{F}_1 + 2\hat{F}_1^2 = 2(\hat{F}_1 - 1)^2 - 1 = 2BS_n(\hat{F}_1, O = 1) - 1. \quad (1.74)$$

De même, si $O = 0$:

$$SPL_n(\hat{F}_1, O = 0) = (1 - \hat{F}_1)^2 + \hat{F}_1^2 - 2(1 - \hat{F}_1) = -1 + 2\hat{F}_1^2 = 2BS_n(\hat{F}_1, O = 0) - 1. \quad (1.75)$$

Pour le *Winkler score*, on rappelle que :

$$WS_n = \begin{cases} -\frac{(1 - \bar{O})^2 - (1 - \hat{F})^2}{\bar{O}^2 \mathbb{1}_{p < \bar{O}} + (1 - \bar{O})^2 \mathbb{1}_{p \geq \bar{O}}} & \text{si } O=1, \\ -\frac{\bar{O}^2 - \hat{F}^2}{\bar{O}^2 \mathbb{1}_{p < \bar{O}} + (1 - \bar{O})^2 \mathbb{1}_{p \geq \bar{O}}} & \text{si } O=0. \end{cases} \quad (1.76)$$

Supposer que $\bar{O} = 0.5$ implique que :

$$WS_n = \begin{cases} -\frac{0.25 - (1 - \hat{F}_1)^2}{0.25} = 4BS_n(\hat{F}_1, O = 1) - 1 & \text{si } O=1, \\ -\frac{0.25 - \hat{F}_1^2}{0.25} = 4BS_n(\hat{F}_1, O = 0) - 1 & \text{si } O=0. \end{cases} \quad (1.77)$$

On a montré ici que le *Brier score*, le *score propre linéaire* et le *Winkler score* dans le cas particulier où la climatologie vaut $\frac{1}{2}$ sont équivalents, au sens de Dawid (2006).

Chapitre 2

Un nouveau modèle de référence climatologique pour les prévisions solaires probabilistes

Résumé

Ce chapitre a pour objectif de compléter le chapitre 1 dans la proposition d'un cadre global d'évaluation de la qualité des prévisions probabilistes. Il a été établi dans le chapitre 0 qu'un tel cadre nécessite un modèle simple à implémenter nommé "benchmark model" dont la qualité sert de référence pour être comparée avec la qualité de nouveaux modèles de prévision. Cette comparaison peut se faire en particulier à l'aide de scores de compétence (cf. partie 1.4.4). Il existe différentes classes de modèles de référence. Nous développons dans ce chapitre la notion de modèle de référence climatologique. Nous présentons une charte de règles régissant les modèles de référence climatologiques. Nous proposons ensuite un nouveau modèle de référence que nous appelons "CSD-CLIM". Celui-ci est ensuite comparé à l'aune des règles proposées au CH-PeEn et à la climatologie naïve, qui sont deux autres modèles de référence climatologiques couramment utilisés par la communauté scientifique. Ce chapitre reprend très largement l'article nommé "A new climatology reference model to benchmark probabilistic solar forecasts" accepté pour publication et qui paraîtra prochainement dans la revue internationale à comité de lecture "Solar Energy". Par rapport à cette publication, la mise en page et certaines notations ont été modifiées afin de les rendre cohérentes avec l'ensemble du travail de thèse.

Sommaire

2.1	Introduction	59
2.2	Required properties for a good climatology benchmark model	61
2.3	Context of the study	62
2.3.1	Data	62
2.3.2	Clear-sky model	63
2.3.3	Generating the predictive cumulative distribution function (CDF)	64
2.3.4	Verification of probabilistic forecasts	65
2.4	Climatology benchmark models for solar probabilistic forecasting	66
2.4.1	The naive climatology (CLIM)	66

2.4.2	The clearsky-dependent Climatology (CSD-CLIM)	67
2.4.3	The Complete-History Persistence Ensemble (CH-PeEn) . . .	70
2.5	Results	70
2.5.1	Compliance with property P2	71
2.5.2	Compliance with property P3	71
2.5.3	Compliance with property P4	71
2.5.4	Compliance with property A1	72
2.5.5	Compliance with property A2	73
2.5.6	Overview	74
2.6	Discussion	75
2.6.1	Discussion on <i>CSD-UNC</i> , the average <i>CRPS</i> obtained by CSD-CLIM	75
2.6.2	Comparison of the binning approaches used by CH-PeEn and CSD-CLIM	76
2.7	Conclusions	79
	Appendices	81
2.A	Sensitivity Analysis related to the number of bins used by the CSD-CLIM binning process	81
2.B	Time-invariance of CSD-CLIM	82
2.C	Reliability assessment using PIT Histograms	83

A new climatology reference model to benchmark probabilistic solar forecasts

Abstract

Probabilistic solar forecasting is becoming a major topic in the solar research community as it provides more information about the uncertainty of the forecast compared to deterministic forecasting. However, to facilitate the adoption of probabilistic forecasts within solar forecasting communities (industry and academic), the definition and the use of standardized best practices are a prerequisite. Among others, there is a need for benchmark models that are able to properly assess the performance of new probabilistic forecasting methods. In this work, we propose a new climatology benchmark model called “CSD-CLIM” (for “Clear-Sky Dependent Climatology”). This new reference model is evaluated against two other climatology benchmark models namely the naive climatology and a well-referenced model in the literature, the CH-PeEn (for Complete History Persistence Ensemble). The verification of compliance with a set of properties that a climatology benchmark model must follow demonstrates that the new CSD-CLIM model outperforms the naive climatology and that it can be a viable alternative to the CH-PeEn model. It is shown that the better performance of CSD-CLIM is due to a specific binning of the historical irradiance data based on the clear-sky irradiance values.

2.1 Introduction

It is now commonly accepted that solar forecasting is a cost-effective way to increase the share of solar energy in the electrical grid (Pierro et al., 2019). Recently, there has been a growing interest in solar probabilistic forecasting (Van der Meer et al., 2018). Indeed, contrary to deterministic forecasts, probabilistic forecasts of intrinsically highly variable weather predictands like wind or solar bring more value to the grid operator as demonstrated by Zhu et al. (2002) or Buizza (2008).

However, as for deterministic forecasts, a wide adoption of probabilistic forecasts in the solar forecasting communities (industry and academic) requires a set of best practices. For instance, one can cite first the existence of a specific framework for verifying the quality of solar probabilistic forecasts. Raising concerns about the verification of probabilistic forecasts and notably the use of improper scores to measure the performance of the probabilistic methods, Lauret et al. (2019) have recently recommended a set of diagnostic tools and numerical scoring rules like the Continuous Ranked Probability Score (*CRPS*) to assess the quality of solar probabilistic forecasts. The $CRPS_n$ measures the integral of the quadratic difference between the cumulative distribution function (“CDF”) of the forecast and the CDF of the corresponding observation. Averaging the $CRPS_n$ of numerous forecast/observation pairs forms the \overline{CRPS} or “*CRPS*” (see section 2.3.4 for details).

A second point is related to the use of well-accepted reference models to fairly benchmark any new proposed forecasting methods on preferably standardized datasets (Yang et al., 2020a). In the realm of solar probabilistic forecasts, a reference model called the Persistence ensemble (PeEn) model (Alessandrini et al., 2015) is routinely proposed to benchmark new probabilistic models (David et al., 2016). Unfortunately, Doubleday et al. (2020) and Yang (2019) noted a wide spectrum of implementations of the PeEn model in

the literature. Besides, Yang (2019) emphasized the need for universal benchmark models whose implementation must depend only on the data at a particular site. This is why Yang (2019) proposed a universal benchmarking model, the Complete History-Persistence ensemble (CH-PeEn). As stated by Yang (2019), the CH-PeEn model constitutes a consistent baseline model for assessing the skill of a forecasting method. The requirement of such benchmarks models has been also highlighted in Doubleday et al. (2020). In their work, the authors compared ten variants of six reference models, and the class of climatology reference models such as naive (no-skill) classical climatology and the CH-PeEn were implemented.

Similarly to skill scores used in the case of deterministic forecasts (Yang et al., 2020b), skill scores like the “*CRPS skill score*”, or “*CRPSS*” can be used to gauge the performance of a new probabilistic forecasting model against a reference easy-to-implement method. However, as noted by Yang (2019), different implementations of the benchmark model can hamper the interpretability of the skill score. Therefore, the computation of these skill scores should be done using a universal well-accepted benchmark model. Such a practice will promote a fair evaluation of probabilistic forecasting techniques.

Let us stress here the importance of using skills scores. Indeed, a score like the average *CRPS* obtained by a forecasting method is not in itself a measure of the skills of the forecast as the score strongly depends on the sky conditions of the considered location. For example, Alessandrini et al. (2015) pointed out that meteorological conditions of a site impact the quality of solar probabilistic forecasts. Hence, the *CRPS* must be compared with the *CRPS* of the reference model. The latter is expected to reflect the difficulty of forecasting at a particular site or equivalently to quantify the predictability of the solar irradiance at that specific location.

Following the need of easy-to-implement (naive) and universal reference models, it has been shown in Murphy (1973) that one component of the *CRPS* called the uncertainty corresponds to the *CRPS* of the climatology. The score of this naive climatology is only sensitive to the observations variability and therefore, for a given location and temporal resolution of the data, does not depend on any other kind of parameters. Thus, one way to avoid a *CRPSS* that depends on the implementation of the reference model is to use the uncertainty part of the *CRPS* as the baseline value. Moreover, it must be noted that, for meteorologists, the baseline model for computing skill scores is usually the climatology - see for instance Cusack and Arribas (2008) or Binter (2012).

However, while appealing, we will show in this work that the naive climatology is not the best candidate for being a reference model for the particular case of solar irradiance forecasting. Indeed, the raw GHI time series exhibit specific diurnal and seasonal patterns which are not taken into account by the naive climatology model. Consequently, these deterministic patterns increase the average *CRPS* of this benchmark model (denoted hereafter *UNC*)¹ Hence, it would be desirable to design a benchmark model which does not suffer from this issue.

In this work, we propose to take advantage of the clear-sky irradiance in order to compute a new climatology reference model called the clear-sky dependent climatology or “*CSD-CLIM*”. Unlike the CH-PeEn reference model which relies on hour-dependent predictive distributions, the CSD-CLIM model makes use of a binning process of the clear-sky irradiance to compute its *CRPS*. Additionally, it will be shown that, unlike CH-PeEn, the score of the CSD-CLIM can be directly computed from the historical data at

¹Note that *UNC* is also the uncertainty component of the *CRPS*.

hand without needing to first form the predictive distribution and then calculate the score on the historical dataset.

We will show also that this new model improves on the notion of universal benchmarking. Besides, we will demonstrate that the *CRPS* of CSD-CLIM and CH-PeEn can reflect the difficulty of forecasting at a particular site.

This paper is structured as follows. Section 2.2 discusses the required properties that a good benchmark model should exhibit. Section 2.3 presents the context of the study and in particular, the data and sites used to assess the performance of the different climatology benchmark models. Section 2.4 details the benchmark models while section 2.5 verifies the compliance of the reference models with the required properties. A discussion related to the score obtained by the CSD-CLIM model and a detailed comparison of the methodologies pertaining to the CSD-CLIM and CH-PeEn models is conducted in section 2.6. Finally, section 2.7 will present our conclusions.

2.2 Required properties for a good climatology benchmark model

In this work, we focus on the class of climatology reference models which can be used to benchmark either intra-hour or hourly forecasts according to the terminology used by Doubleday et al. (2020). Particular attention must be paid to the selection of a benchmark model and the amount of information used to feed it. Since a benchmark model is mainly used to calculate a skill score of a new forecasting method², the following (at least) two questions³ related to the formulation of a skill score must be addressed:

1. What information should be given by a skill score ?
2. What score should correspond to the “0” skill score?

This could be subject to discussion, but in our opinion, the purpose of a skill score should be to indicate which part of the information given by the new forecast is naive (i.e. captured by the benchmark model), and to what extent it provides valuable extra-information (which should be credited to the particular skill of the forecast).

Furthermore, when dealing with climatology benchmarks, we propose that the “0” skill score should be defined by the best possible exploitation of all information derived from historical observations (i.e. the climatology). Thus, all the historical data should be exploitable by the benchmark model. Conversely, all extra-information (from meteorological data, satellite observation, etc.) treated by the new forecasting method and the potentially associated better performance should be credited to the merit of the forecast.

We hypothesize that the preceding propositions hold and we propose in this section to establish a set of required properties that climatology benchmark models should meet. This set is intended to reflect the discussion conducted above about the role of a climatology benchmark model. Doubleday et al. (2020) and Yang (2019) have already highlighted four properties that we retain in this study and form what we call here “strong” conditions for climatology benchmark. Table 2.2.1 lists these four attributes. We propose

²For a negatively oriented score like the *CRPS* for instance, the value of the skill score ranges from $-\infty$ for the worst forecast to 1 for a perfect forecast.

³Other questions could be considered e.g. “Is the skill score universal and applicable on all datasets? (by some measure)”

also in Table 2.2.1 two additional properties that we believe a climatology benchmark model should have. These additional properties are used to further classify the different climatology reference models that satisfy the “strong” conditions.

Code	Property
P1	The benchmark model should be easy-to-implement
P2	The implementation of the model must depend exclusively on the historical data at hand (and not on any other kind of parameters such as number of past measurements, forecast horizon/lead time)
P3	The score (average <i>CRPS</i>) of the model (for a specific location and time resolution of the data) must be unique (or near unique) irrespective of the period or length of the period used to compute the score (“Time-invariance property”)
P4	The model should verify the statistical consistency of the naive climatology (i.e. a perfect reliability when compared to new observations)
A1	The quality of the benchmark model should be as high as possible (i.e. the related average <i>CRPS</i> should be as low as possible)
A2	The score of the benchmark model should reflect the difficulty of forecasting at a particular location

Table 2.2.1: Properties of a good climatology benchmark model. “P” refers to strong conditions for climatology benchmark. These properties were already mentioned in the literature. We propose in this study to add the properties denoted by “A” that permit to discriminate further the “strong” climatological benchmark models.

Some of these properties, like P2 and P4, are closely related to the climatological nature of the class of benchmark models considered here. Consequently, this study excludes the following models :

- PeEn (“Persistence Ensemble”) developed by Alessandrini et al. (2015) and used for example in David et al. (2016), Lauret et al. (2017), and Pedro et al. (2018),
- raw ensemble forecasts used in Golestaneh et al. (2016b) and Thorey et al. (2018),
- other benchmark models based on simple post-processing of raw ensemble forecasts.

However, and as recommended by Doubleday et al. (2020), in addition to the climatology, other type of models such as the raw ensemble must be used to benchmark hourly forecasts.

2.3 Context of the study

2.3.1 Data

A selection of 20 sites serves as support for the comparison of the different benchmark models. This choice was made by trying to keep the widest possible spectrum in terms

of different sky conditions and locations around the world. The vast majority of data was chosen from BSRN (“Baseline Surface Radiation Network”, <https://bsrn.awi.de/>) collection data in order to minimize the differences in data acquisition and data validation between sites. The remaining data comes from previous works dedicated to state-of-the-art of solar forecasting (David et al., 2016; David et al., 2018; Le Gal La Salle et al., 2020). The complete list of sites is given in Table 2.3.1. The observation data considered are global horizontal irradiance (GHI) measurements. As BSRN data is submitted to strict data quality checks, we consider that the quality of data is good. Note that the quality checks of all BSRN data are available in BSRN website. For quality checks of data which is not part of BSRN, please refer to Le Gal La Salle et al. (2020).

	Provider	Data	Longitude	Latitude	Climate	Acronym
Saint-Pierre	PIMENT	2012-2013	55.49	-21.34	Insular tropic	SPI
Hawaii	NREL	2010-2011	-158.08	21.31	Insular tropic	HAW
Desert Rock	BSRN	2012-2013	-116.03	36.62	Desert	DRO
Fort Peck	BSRN	2012-2013	-106.50	48.00	Continent	FPE
Fouillol	LARGE	2010-2011	-61.52	16.22	Insular tropic	FOU
Payerne	BSRN	2017-2018	6.94	46.82	Mountainous	PAY
Palaiseau	BSRN	2016-2017	2.21	48.71	Temperate	PAL
Toravere	BSRN	2017-2018	26.46	58.25	Temperate	TOR
Adelaïde	BSRN	2016-2017	138.51	-35.00	Desert	ADE
Tiruvallur	BSRN	2015-2016	79.97	13.09	Monsoon	TIR
Sioux Falls	BSRN	2017-2018	-96.62	43.73	Continent	SXF
Nauru Islands	BSRN	2011-2012	166.92	-0.52	Insular tropic	NAU
Marshall Islands	BSRN	2014-2015	167.73	8.72	Insular tropic	MAR
Barrow	BSRN	2015-2016	-156.61	71.32	Arctic	BAR
Cocos Islands	BSRN	2017-2018	96.84	-12.19	Insular tropic	COC
Manus Islands	BSRN	2011-2012	147.43	-2.06	Insular tropic	MAN
Tenerife	BSRN	2017-2018	-16.5	28.31	Temperate	TEN
Minamitorishima	BSRN	2017-2018	153.98	24.29	Insular	MIN
Bermuda Islands	BSRN	2011-2012	-64.67	32.27	Insular	BER
Langley	BSRN	2018-2019	-76.39	37.10	Temperate	LAN

Table 2.3.1: Characteristics of sites used for the study.

2.3.2 Clear-sky model

The building of the CH-PeEn and CSD-CLIM benchmark models (see section 2.4) requires a clear-sky model. As demonstrated by Yang (2019) who compared two clear-sky models in the assessment of the performance of the CH-PeEn model, the choice of the clear-sky model is of primary importance (See also Yang (2020)).

Among the different clear-sky models that can be found in the literature, one can cite the Bird Model (Bird and Hulstrom, 1981), the Ineichen-Perez model (Ineichen and Perez, 2002) or the McClear model (Lefèvre et al., 2013). In this work, and following the work of Yang (2019), we have selected the McClear clear-sky model. McClear Clear-sky GHI estimates are publicly available at 1-min resolution from the CAMS (Copernicus Atmosphere Monitoring Service) McClear Service (Copernicus Atmosphere Monitoring Service, n.d.).

2.3.3 Generating the predictive cumulative distribution function (CDF)

The computation of the *CRPS* requires the building of the predictive cumulative distribution function (CDF) $\hat{\mathcal{F}}$ (see Equation 2.2). In this section, we describe briefly how these forecast CDFs can be generated.

Generation of climatology predictive distributions

Every climatology models used in this study are built on the idea of combining past observations to create predictive distributions. For instance, the naive climatology forecast is an empirical CDF based on long period of N_A historical sorted measurements (O_1, O_2, \dots, O_{N_A}). Here, to generate the predictive CDF for the benchmark models, we implement the classical approach (Lauret et al., 2019) that consists in building a piecewise constant function with a jump probability of $\frac{1}{N_A}$ at each O_i and null probabilities for events outside the set of historical measurements. The predictive CDF is given by :

$$\hat{\mathcal{F}}(x) = \frac{1}{N_A} \sum_{i=1}^{N_A} \mathbb{1}_{\{x \geq O_i\}}, \quad (2.1)$$

where $\mathbb{1}_{\{u\}}$ is the indicator function which has the value of 1 if its argument u is true and 0 otherwise. Note that in the case of the CH-PeEn model, predictive CDFs are built with a training set of historical ordered measurements for each hour of the day.

Generation of the day-ahead LQR probabilistic forecasts

Property A2 states that the average *CRPS* of the benchmark model should be a proxy for judging a priori the quality of a probabilistic forecast. In order to be able to evaluate the proposed benchmark models regarding this property, we generate day-ahead probabilistic forecasts for the different sites listed in Table 2.3.1.

The day-ahead GHI ensemble forecast has been provided by the European Centre of Medium-Range Weather Forecasts (ECMWF). This ensemble forecast also called EPS (for ensemble prediction system) is constituted of 51 members : one unperturbed member (control member) and 50 perturbed members. The temporal resolution is of 3 hours and the spatial resolution is of 0.2° in both longitude and latitude. Consequently, 3h GHI times series recorded on-site are compared with the nearest ECWFMF pixel.

Numerous studies (see for instance Vannitsem et al. (2018a)) have shown that CDFs drawn from raw EPS ensemble are statistically unreliable, meaning that the probability assigned to an event is not consistent with the observations (Hamill and Colucci, 1997). Hence, the use of calibration models (i.e. techniques which improve the reliability of raw ensemble forecasts) to create CDFs from ensemble forecasts is a common practice. The interested reader is referred to Gneiting et al. (2005) or Le Gal La Salle et al. (2020) for details regarding the implementation of calibration techniques.

In this study, we propose to use a state-of-the-art non-parametric and very flexible calibration method, the Linear Quantile Regression (“LQR”) technique. The LQR method is depicted at length in Le Gal La Salle et al. (2020).

2.3.4 Verification of probabilistic forecasts

In the verification framework proposed by Lauret et al. (2019), the authors recommend the computation of a proper score like the Continuous Ranked Probability Score (*CRPS*) to evaluate the overall quality of a probabilistic forecast. We will recall here its mathematical definition.

The $CRPS_n$ is a measure of the distance between the forecast CDF $\hat{\mathcal{F}}_n$ and the CDF associated with the measurement O_n (Hersbach, 2000) :

$$CRPS = \int_{Clim_{\min}}^{Clim_{\max}} \left[\hat{\mathcal{F}}_n(x) - \mathcal{O}_n(x) \right]^2 dx. \quad (2.2)$$

$\hat{\mathcal{F}}_n(x)$ is the predictive CDF and $\mathcal{O}(x)$ is the cumulative distribution given by the Heaviside (or step) function $H(x - O)$, which is zero if $x < O$ and one if $x \geq O$. The $CRPS_n$ can be averaged over a set of N forecast/observation pairs, which defines the \overline{CRPS} :

$$\overline{CRPS} = \frac{1}{N} \sum_{n=1}^N CRPS_n. \quad (2.3)$$

hereinafter simply denoted as *CRPS*. The *CRPS* is negatively oriented (smaller values indicate a better forecast). Like the *Brier score*, the *CRPS* is a proper score and can be decomposed as follows :

$$CRPS = REL - RES + UNC, \quad (2.4)$$

where *REL*, *RES* and *UNC* are respectively the reliability part, the resolution part and the uncertainty part of the *CRPS*.

Reliability is an indication of the statistical consistency between the forecasts and the observations while resolution indicates how far the observations are discriminated from the climatological mean by the forecasts. Finally, the uncertainty term depends on the variability of the observations and will be further developed in section 2.4.1. As mentioned in the introduction, the uncertainty term corresponds theoretically to the *CRPS* of the naive climatology. Indeed, if we assume an infinite historical time series, the reliability is perfect (i.e. $REL = 0$) and the resolution is null (i.e. $RES = 0$). More details regarding the calculation of the different components of the *CRPS* can be found in Lauret et al. (2019).

Note that, in order to assess the reliability and sharpness properties of the different benchmark models, Doubleday et al. (2020) used visual diagnostic tools like reliability and sharpness diagrams. In this work, we rely on the decomposition of the *CRPS* in order to obtain a quantitative measure of these two attributes. However, as mentioned by Candille and Talagrand (2005), the *CRPS* decomposition may suffer from some limitations, for example when the number N of forecast/observation pairs is small.

Finally, and as mentioned in the introduction, the purpose of the benchmark models is mainly to be used as references in the calculation of skill scores. A skill score is the level of improvement of a forecasting model over the reference model. For example, the *CRPS* skill score reads as :

$$CRPSS = 1 - \frac{CRPS_{model}}{CRPS_{reference}}. \quad (2.5)$$

2.4 Climatology benchmark models for solar probabilistic forecasting

In this section, three reference models, the naive climatology (CLIM), the clear-sky dependent climatology (CSD-CLIM) and the Complete-History Persistence Ensemble (CH-PeEn) are presented and their possible pros and cons are discussed.

2.4.1 The naive climatology (CLIM)

The climatology refers to the ensemble of all the observed values of a weather variable over a long period of time. The predictive CDF created from the aggregation of all these past observations forms the climatological predictive distribution. We denote hereafter the corresponding model as the naive climatology model (CLIM). The *CRPS* of the CLIM model can be computed either by using Equation 2.2 or by computing the uncertainty part of the *CRPS* which reads as (Todter and Ahrens, 2012)

$$UNC = \int_{Clim_{min}}^{Clim_{max}} UNC_{BS}(x) dx. \quad (2.6)$$

$UNC_{BS}(x)$ is the uncertainty relative to the *Brier score* for a fixed level of irradiance x and $Clim_{max}$ and $Clim_{min}$ are the extremum possible values of irradiance (also called climatological bounds) proposed by Yang (2019) and Long and Shi (2008).

Let O be the measurement of the predictand (here the GHI). For a fixed level x of GHI, $UNC_{BS}(x)$ is defined as

$$UNC_{BS}(x) = \bar{O}(x) \left(1 - \bar{O}(x) \right), \quad (2.7)$$

where \bar{O} is the frequency with which O is lower or equal to x i.e.

$$\bar{O}(x) = \frac{\mathbb{1}_{\{O \leq x\}}}{N}, \quad (2.8)$$

and N is the number of historical observations and $\mathbb{1}_{\{u\}}$ the indicator function.

Note that Equation 2.6 allows a graphical representation of UNC which will be extensively used in this study. For each level of GHI x , a point equal to $UNC_{BS}(x)$ is plotted, and UNC is given by the area under the curve created by this set of points. Such a representation is given in Figure 2.4.1.

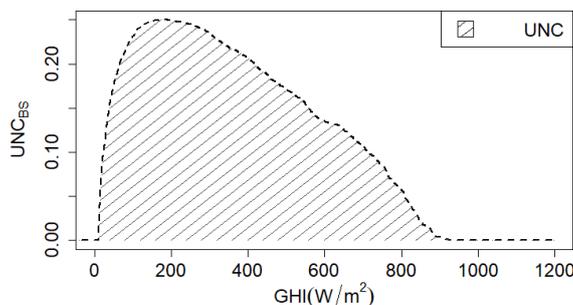


Figure 2.4.1: Uncertainty of the *Brier score* in relation with the level of GHI for the site of Payerne. The uncertainty part of the *CRPS* (UNC) is given by the shaded area.

2.4.2 The clearsky-dependent Climatology (CSD-CLIM)

In the meteorological community, the naive climatology depicted above is used as a reference model for the calculation of skill scores. However, it does not account for the strong time dependence experienced by a variable such as GHI (daily and seasonal patterns).

It is well-known that the irradiance can be decomposed as

$$\text{GHI} = \text{GHI}_{\text{CS}} * k_t^*, \quad (2.9)$$

where GHI_{CS} stands for the clear-sky irradiance and represents the irradiance if no cloud cover is observed, and is given by a chosen clear-sky model (see section 2.3.2). The clear-sky irradiance is fundamentally time-dependent and follows strong temporal patterns.

k_t^* is the clear-sky index and ranges theoretically between 0 to 1⁴ and represents the share of GHI_{CS} which is lost due to cloud cover. Once a clear-sky model is selected, Equation 2.9 shows that forecasting k_t^* and forecasting irradiance are equivalent tasks. Indeed, GHI_{CS} is deterministically estimated by the clear-sky model⁵. However, this is not taken into account by the naive climatology, which only reflects the variability of the GHI. For example, a site which always experiences clear sky conditions (i.e. $k_t^* = 1$) will exhibit a high uncertainty *UNC*, even though the forecast is not difficult (knowing that $k_t^* = 1$). Also, for various cases (e.g. early morning, late evening), the clear-sky model alone, which is not the uncertain part of the forecast, permits to determine that GHI is limited to low-irradiance values. These considerations lead to the idea that k_t^* and the clear-sky model should play a key role in the building of a benchmark model.

This is why we propose here a new climatology reference model that exploits the deterministic variations of the clear-sky irradiance captured by the clear-sky model. We propose to bin the data according to the clear-sky irradiance values and to call this baseline model the clearsky-dependent climatology (“CSD-CLIM”). The main goal of such a model is to discriminate the situations that the clear-sky model alone permits to separate. The implementation of CSD-CLIM and the computation of its related score are depicted below.

A number of bins of clear-sky irradiances (Nb) is chosen. For each bin, CSD-CLIM computes a forecast CDF in a similar manner as the climatology does (see Equation 2.1), using only the historical GHI training data belonging to the bin. Nb predictive CDFs are therefore generated. The *CRPS* of CSD-CLIM on a test dataset can be classically computed using Equation 2.3. More precisely, to compute the $CRPS_n$ of a forecast/observation pair, the level of the corresponding clear-sky irradiance is extracted and the bin related to this level is selected. The CDF associated with this specific bin is used as the forecast in Equation 2.2.

Alternatively, the score (i.e. the *CRPS*) of this new benchmark model can be directly calculated from the test dataset partitioned into Nb bins using

$$\text{CSD-UNC} = \sum_i^{Nb} f_i * \text{UNC}^i. \quad (2.10)$$

For each bin i of clear-sky irradiance values (see Table 2.4.1), we calculate UNC^i as defined by Equation 2.6 and the frequency f_i represents the relative frequency of each bin

⁴In practice, cloud enhancement events (i.e. multi-reflections of the sun beams by the clouds) can produce over-irradiance with clear-sky indices superior to 1.

⁵Note that Cros et al. (2013) highlighted the closeness of results issued from commonly used clear-sky models.

in the clear-sky model. Considering the close relationship between this new model and the naive climatology, we propose to call the *CRPS* of this benchmark model the “clearsky-dependent uncertainty” or “*CSD-UNC*”, just as the score of the naive climatology is called uncertainty (*UNC*) in the literature (see for instance Todter and Ahrens (2012)).

We recommend to compute equally-sized bins from GHI_{MIN} to GHI_{MAX} . The choice of the number of bins Nb could be questioned, and is discussed in Appendix 2.A. From this analysis, we point out that $Nb = 30$ could be selected as a standard choice for hourly forecasts. In practice, other choices will give similar results provided that Nb is reasonably large. In this study, we chose a number of $Nb = 30$ bins. An example of the binning process for Desert Rock is shown in Table 2.4.1.

bin i (W/m^2)	0-40	40-80	80-120	120-160	160-200	200-240
relative frequency f_i	0.06	0.07	0.06	0.05	0.04	0.04
bin i (W/m^2)	240-280	280-320	320-360	360-400	400-440	440-480
relative frequency f_i	0.04	0.04	0.02	0.03	0.02	0.02
bin i (W/m^2)	480-520	520-560	560-600	600-640	640-680	680-720
relative frequency f_i	0.02	0.05	0.05	0.04	0.04	0.05
bin i (W/m^2)	720-760	760-800	800-840	840-880	880-920	920-960
relative frequency f_i	0.06	0.04	0.04	0.02	0.02	0.02
bin i (W/m^2)	960-1000	1000-1040	1040-1080	1080-1120	1120-1160	1160-1200
relative frequency f_i	0.03	0.02	0.01	0.00	0.00	0.00

Table 2.4.1: Binning process for CSD-CLIM in Desert Rock. The relative frequency f_i is the number of clear-sky irradiance values in the bin out of the total number of irradiances values.

Let us stress here that there is no need to form a predictive CDF and compute its *CRPS* with Equation 2.3. Instead, Equation 2.10 together with Equation 2.6 fully determine the score *CSD-UNC* of the proposed new reference model CSD-CLIM. Hence, based on the historical data at hand (property P2 verified), we can argue that the CSD-CLIM model meets the “easy-to-implement” property P1. Furthermore, this ability to be computed without forming a CDF allows one to consider *CSD-UNC* either as the *CRPS* of a prediction model, CSD-CLIM, or as an intrinsic characterization of a particular data set.

Graphically, we propose to represent this *CSD-UNC* by stacking all the UNC^i scaled by their frequencies. As an illustration, Figure 2.4.2 shows the stacked 30 UNC^i . The sum of these areas amounts to the *CSD-UNC*. It becomes then possible to graphically compare *CSD-UNC* and the classical *UNC* defined in section 2.4.1. Such a comparison for two sites (Marshall Island and Desert Rock) is done in Figure 2.4.3.

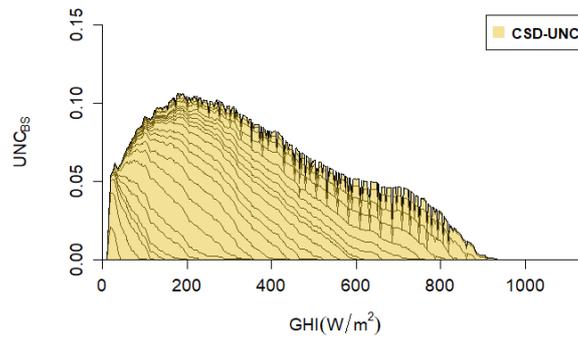


Figure 2.4.2: Construction of *CSD-UNC* for Payerne. The sum of the 30 colored areas gives *CSD-UNC*.

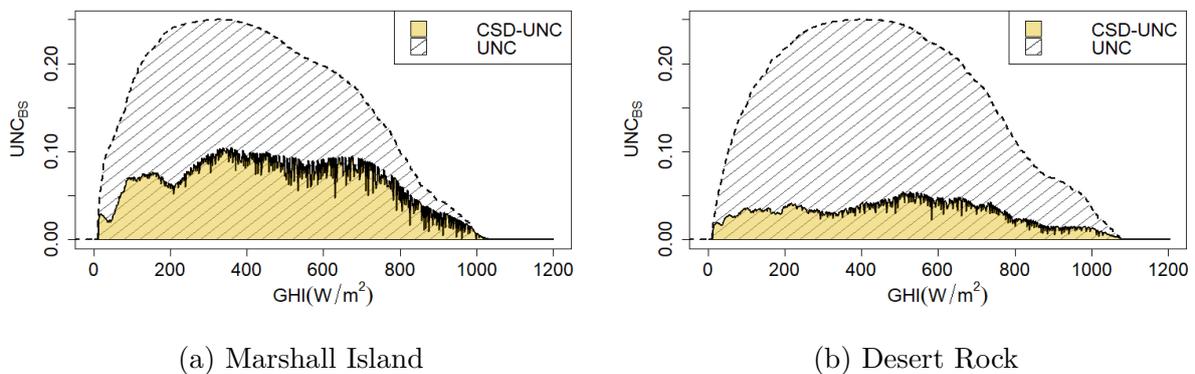


Figure 2.4.3: Comparison between *UNC* and *CSD-UNC*.

As shown by Figure 2.4.3, a substantial difference can exist between *UNC* and *CSD-UNC* (and by extension between the two reference models CLIM and CSD-CLIM). While for Marshall Island, *CSD-UNC* represents approximately 50% of *UNC*, the reduction is more drastic in the case of Desert Rock. This discrepancy can be explained with the histograms of k_t^* presented in Figure 2.4.4 which reveals that k_t^* is both lower and much more variable in Marshall Island than in Desert Rock. This has two consequences:

1. *UNC* is more important in Desert Rock, mainly because k_t^* is in average higher leading to a distribution of higher values of GHI (see Figure 2.4.3),
2. The forecasting task, which essentially consists in guessing the most probable value of k_t^* , is much more difficult in Marshall Island.

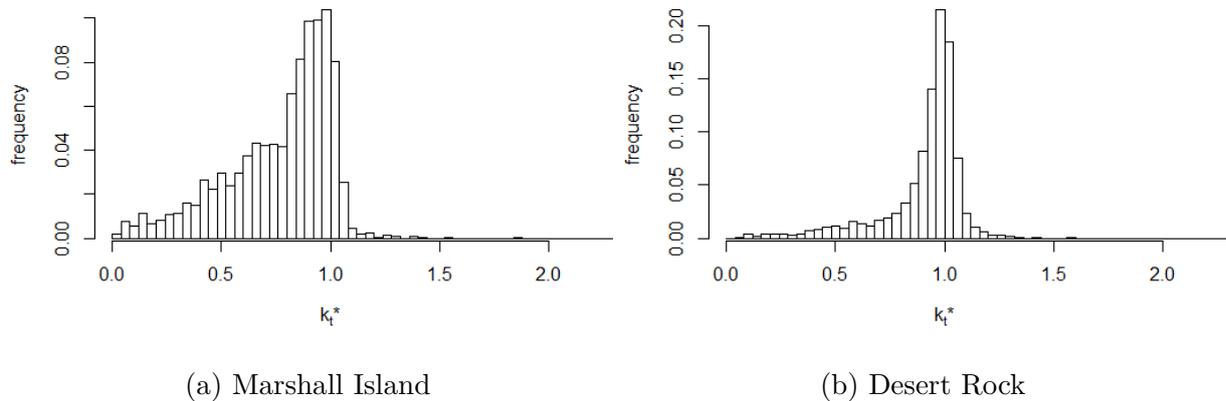


Figure 2.4.4: Histograms of k_t^* for Marshall Island and Desert Rock.

2.4.3 The Complete-History Persistence Ensemble (CH-PeEn)

By pointing out the deficiencies of the benchmark model widely used by the solar forecasting community namely the Persistence ensemble PeEn (Alessandrini et al. (2015), David et al. (2016)), Yang (2019) proposed a new reference model i.e. the Complete-History Persistence Ensemble (CH-PeEn). Unlike the PeEn model whose implementation depends heavily on the number of previous measurements used to build the predictive CDF, the proposed CH-PeEn baseline model uses the entire set of past measurements to form predictive distributions conditioned by each hour of the day. As discussed by Yang (2019), such a benchmark model like CH-PeEn will have a near unique *CRPS* and therefore will ease the interpretation of skill scores.

Regarding more precisely the implementation of the CH-PeEn model, based on the entire history of available data, the empirical CDF of the clear-sky indices k_t^* for each hour is formed. To compute the $CRPS_n$ (see Equation 2.2) of the CH-PeEn model, the predictive GHI distribution is obtained by multiplying the empirical set of k_t^* of a given hour by the clear-sky irradiance value at this specific forecast hour. As a set of k_t^* is independently created for each hour, the resulting predictive CDF is strongly dependent on the local hour. It must be stressed also that, unlike *CSD-UNC* whose computation does not require the construction of predictive GHI distributions, the derivation of CH-PeEn implies the generation of forecasts and the assessment of the $CRPS_n$ of these forecasts.

2.5 Results

In this section, we verify the compliance of the different benchmark models with the properties listed in Table 2.2.1. It must be noted that, by construction, the 3 climatology benchmark models meet property P1. The goal of this section is then to find whether the models follow P2 and P3, and to evaluate the performance of the benchmark models in the light of properties P4, A1 and A2.

2.5.1 Compliance with property P2

As claimed by P2 in Table 2.2.1, the implementation of a good benchmark model should require a minimum number of parameters, for both simplicity of calculation and universality of results. The naive climatology is the only model which, by construction, indisputably fully respects this criteria. On the contrary, the other models presented here use some additional parameters.

CH-PeEn needs the time of the day to bin the data and a clear-sky model to compute the clear-sky indices k_t^* . The time of the day is not strictly speaking an additional parameter, because it does not depend on a disputable model. The time resolution of the bin could influence the results (note that this is also true for the naive climatology), but should be governed by the available data, and thus should depend only on the data at hand, as stated in P2. The choice of the clear-sky model could be more problematic, because different models could provide different results. However, it seems reasonable to accept this compromise given the benefits of using a clear-sky model.

CSD-CLIM uses also a clear-sky model to bin the data. The same argument used to justify this usage for CH-PeEn is also valid for CSD-CLIM. Besides, *CSD-UNC* could also depend on the chosen number of bins Nb , and therefore rigorously does not fully meet property P2. We justify in Appendix 2.B that the number of bins has little impact on the final results, as long as Nb is chosen reasonably large. This condition being fulfilled, it can be argued that, in practice, CSD-CLIM nearly complies with P2.

2.5.2 Compliance with property P3

Property P3 states that all benchmark models must be time-invariant i.e. that their resulting *CRPS* (for a specific location and time resolution of the data) must be unique or near unique. The verification of the time invariance of the CSD-CLIM model is detailed in Appendix 2.B. As demonstrated in it, we can conclude that the CSD-CLIM produces a near unique score regardless of the period or the length of the historical data used to compute it. The CH-PeEn is also time-invariant, as demonstrated by Yang (2019). Furthermore, it is well-known that the naive climatology is time-invariant, as soon as the amount of data considered for its computation is sufficiently large.

2.5.3 Compliance with property P4

Property P4 emphasizes the importance of reliability of the benchmark model. We recall that a climatology benchmark model should possess the same statistical consistency as the naive climatology and therefore should exhibit a reliability component as close as possible to zero. In addition, while respecting the statistical consistency property, any other benchmark model should beat the naive CLIM model in terms of resolution.

As mentioned above, contrary to Doubleday et al. (2020) who used visual diagnostic tools like reliability and sharpness diagrams in order to assess the two important attributes of a forecasting scheme i.e reliability and resolution, we prefer here to rely on the quantitative decomposition of the *CRPS*. Table 2.5.1 details the decomposition of the *CRPS* into reliability and resolution obtained by the 3 reference models. For each site, the three models are trained on the first year and tested on the second year (see Table 2.3.1). Results given in Table 2.5.1 are related to the testing period.

	Reliability (W/m^2)									
site	SPI	HAW	DRO	FPE	FOU	PAY	PAL	TOR	ADE	TIR
CSD-CLIM	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
CH-PeEn	7.9	8.0	6.0	6.3	5.1	10.0	6.6	8.8	7.0	5.3
CLIM	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0
site	SXF	NAU	MAR	BAR	COC	MAN	TEN	MIN	BER	LAN
CSD-CLIM	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
CH-PeEn	5.3	4.3	5.0	3.7	5.8	4.8	6.0	4.2	6.0	5.5
CLIM	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.1
	Resolution (W/m^2)									
site	SPI	HAW	DRO	FPE	FOU	PAY	PAL	TOR	ADE	TIR
CSD-CLIM	113.2	107.1	140.4	89.1	63.9	85.0	72.0	66.7	104.0	110.2
CH-PeEn	119.3	109.4	145.0	92.8	67.7	88.3	74.3	70.4	109.0	113.1
CLIM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
site	SXF	NAU	MAR	BAR	COC	MAN	TEN	MIN	BER	LAN
CSD-CLIM	74.9	119.7	95.2	52.5	105.3	83.1	164.4	126.5	94.7	89.6
CH-PeEn	77.8	122.0	97.9	55.3	109.2	86.5	170.5	130.0	98.5	92.8
CLIM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 2.5.1: Decomposition of the *CRPS* for the 3 tested benchmark models and all sites.

As expected, the reliability and resolution components of the naive climatology (CLIM) are zero or near zero. It appears also that while retaining the statistical consistency feature, the CSD-CLIM model improves on the resolution of the basic CLIM model. Conversely, it can be noted that the CH-PeEn exhibits a slightly higher resolution than the CSD-CLIM model but at the expense of a degradation in reliability. In order to strengthen the results presented here, and to avoid the possibility that the reliability could be misjudged due to the biases noted by Candille and Talagrand (2005), who warned on the importance of computing the decomposition of the *CRPS* on large datasets, the reliabilities of the three models for Payerne, Sioux Falls, Tenerife and Bermuda island are also assessed through PIT histograms in Appendix 2.C. The results obtained for these four sites give a good indication of what can be obtained for the 20 sites. As mentioned in Appendix 2.C, the reliability assessments with the help of PIT histograms seem to corroborate the results provided by the decomposition of the *CRPS*.

2.5.4 Compliance with property A1

Additional property A1 states that the benchmark models should obtain the best possible quality using only historical data. The overall quality of the reference forecasts is measured here by the *CRPS*. Table 2.5.2 gives the *CRPS* of the 3 benchmark models for all the sites listed in Table 2.3.1.

Model	SPI	HAW	DRO	FPE	FOU
LQR forecast	51.36	46.82	25.58	39.37	77.89
CLIM	172.7	157.8	175.0	137.8	140.1
CSD-CLIM	59.5	50.7	34.6	48.7	76.2
CH-PeEn	61.2	56.2	35.6	51.2	77.4
Model	PAY	PAL	TOR	ADE	TIR
LQR forecast	40.61	36.14	30.95	41.40	42.49
CLIM	143.4	126.4	116.5	162.4	161.0
CSD-CLIM	58.3	53.5	49.8	48.4	50.8
CH-PeEn	64.8	58.6	54.6	60.4	52.8
Model	SXF	NAU	MAR	BAR	COC
LQR forecast	39.65	54.42	61.12	29.62	48.40
CLIM	137.4	175.7	167.3	88.8	166.8
CSD-CLIM	62.5	56.0	72.1	36.3	61.5
CH-PeEn	65.0	58.0	74.4	37.1	63.3
Model	MAN	TEN	MIN	BER	LAN
LQR forecast	68.25	40.87	33.09	44.16	35.16
CLIM	162.1	199.6	170.0	152.4	154.4
CSD-CLIM	79.0	35.2	43.5	57.7	64.8
CH-PeEn	80.1	35.2	44.0	59.7	67.3

Table 2.5.2: $CRPS$ (W/m^2) of the LQR forecast (grey) and the 3 benchmark models for all sites.

As shown by Table 2.5.2, for any given considered site, the CSD-CLIM model exhibits the highest overall quality in terms of $CRPS$. The previous decomposition of the $CRPS$ shows that the overall better performance of CSD-CLIM originates from its high reliability. This finding strengthens the assumption that knowledge on clear-sky irradiance decreases the uncertainty⁶ associated with a forecast.

2.5.5 Compliance with property A2

As mentioned in section 2.3.3, in order to verify the compliance of the benchmark models with the additional property A2, we generate day-ahead LQR calibrated forecasts. The assumption of this study is that the score of a benchmark model should be related to the quality of the LQR forecasts (measured here by its $CRPS$). Figure 2.5.1 plots the results of the three benchmark models (x axis) versus the $CRPS$ of the LQR forecasts (y axis). Table 2.5.2 also gives the $CRPS$ of the LQR forecasts.

⁶The word “uncertainty” is not used here to refer to the uncertainty term of the decomposition of the $CRPS$, but to the expected level of variability of the predictand, which a forecast model has to deal with.

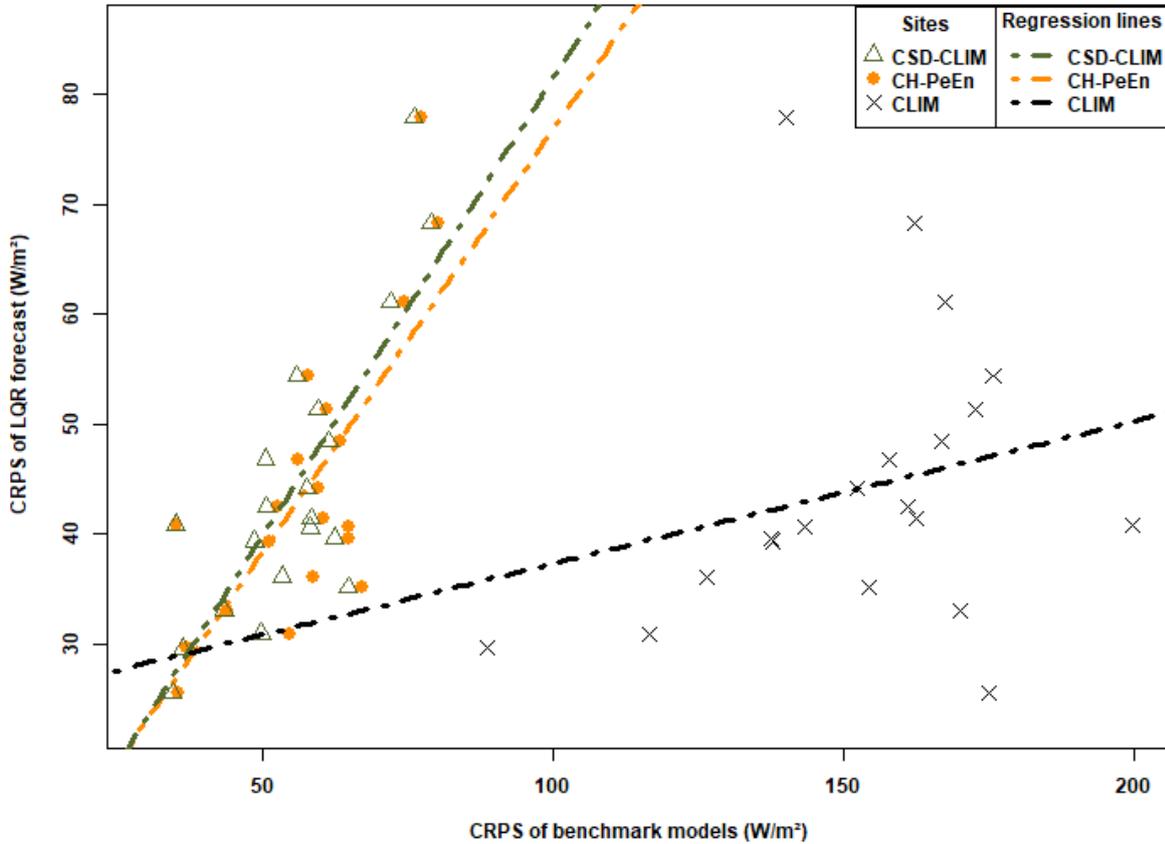


Figure 2.5.1: $CRPS$ of the LQR forecasts vs $CRPS$ of the benchmark models computed for each site of Table 2.3.1.

Figure 2.5.1 clearly shows that, unlike the score of the naive CLIM, the scores of CSD-CLIM and CH-PeEn can be good proxies for judging a priori the quality of a forecast obtained at a particular site. In other words, just like the $RMSE$ score obtained by the clear-sky persistence reference model in case of deterministic forecasts, the score of these two reference models reflects the difficulty of forecasting at a particular location.

To proceed further, we built a linear regression for each considered benchmark model and extracted the coefficient of determination R^2 . We found respectively a R^2 coefficient of 0.06, 0.56 and 0.63 for the CLIM, CH-PeEn and CSD-CLIM. Furthermore, it must be noted that a similar ranking has been established for other types of calibrated forecasting models described in Le Gal La Salle et al. (2020). Put differently, whatever the forecasting model, the best correlation is always obtained by CSD-CLIM. As a conclusion, we can state that CH-PeEn and CSD-CLIM outperform the naive climatology regarding property A2.

2.5.6 Overview

Let us recall that all the climatology benchmark models discussed here meet the required rules P1, P2 (again, let us stress that, in practice, CSD-CLIM nearly complies with this

property) and P3. Finally, Table 2.5.3 gives an overview of the relative performance of each benchmark model in the light of properties P4, A1 and A2.

Model	P4 Statistical consistency	A1 Overall quality of the model	A2 Proxy indicator for forecast difficulty
CLIM			
CH-PeEn			
CSD-CLIM			

Best Model
 2nd best Model
 Worst Model

Table 2.5.3: Overall qualitative assessment of the benchmark models.

The qualitative results of Table 2.5.3 suggest that CSD-CLIM should be preferred as a reference model. It appears to lead to the best trade-off between all the properties required by a climatology benchmark model (see section 2.2).

2.6 Discussion

2.6.1 Discussion on *CSD-UNC*, the average *CRPS* obtained by CSD-CLIM

An in-depth study was conducted in order to understand what drives the differences in the *CRPS* of the CSD-CLIM (i.e *CSD-UNC*) obtained at the different sites. However, in this section, we restrict the analysis on two specific sites namely Desert Rock (low *CSD-UNC*) and Fouillole (high *CSD-UNC*). For these two sites, in order to understand what drives the differences in *CSD-UNC*, the UNC^i is plotted for the 30 bins of clear-sky irradiances in Figure 2.6.1.

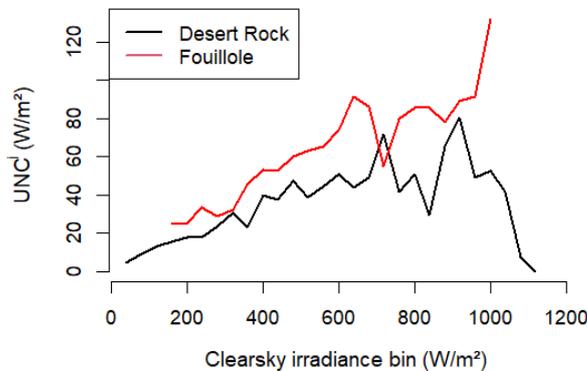


Figure 2.6.1: UNC^i in relation with each bin i of clear-sky irradiance for Desert Rock and Fouillole.

The values of UNC^i generally increases with the level of clear-sky irradiance. Thus, the bins corresponding to high clear-sky irradiances values are responsible for the major part of $CSD-UNC$. Besides, the uncertainty is highest in Fouillole for almost every bin. A closer look on a bin corresponding to high clear-sky irradiances is presented in Figure 2.6.2 and shows from where come these differences.

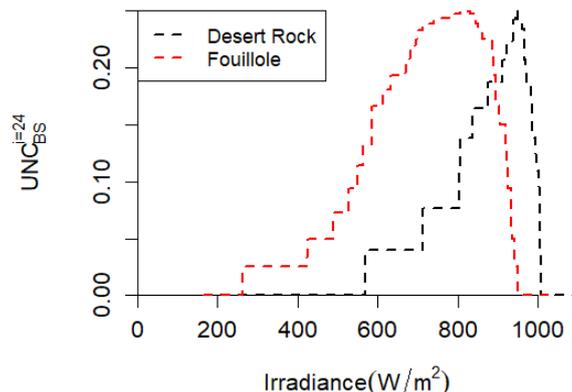


Figure 2.6.2: $UNC^{i=24}$ in Desert Rock and Fouillole for clear-sky irradiances between 900 and 940 W/m^2 (24th bin).

For high clear-sky irradiances, even low GHI measurements can occur in Fouillole, which is not the case at Desert Rock. This enlarges the area under $UNC_{BS}^{i=24}$ for Fouillole, leading to a higher $UNC^{i=24}$, and consequently to a higher $CSD-UNC$.

2.6.2 Comparison of the binning approaches used by CH-PeEn and CSD-CLIM

In terms of $CRPS$, the results obtained with CH-PeEn turned out to be quite comparable with those of CSD-CLIM. This is not surprising since the general idea behind these two models is very close : giving a time-of-the-day dependent image of the uncertainty. The main difference between the two models is related to their approach to binning. The CH-PeEn model groups together all observations made at the exact same hour whereas the CSD-CLIM model proposes to bin the GHI data according to the clear-sky irradiance value. We propose here to use contingency tables in order to better highlight the relative difference in the binning methodology used by each model. Figure 2.6.3 shows such contingency tables for 3 specific sites namely Nauru, Desert Rock and Toravere. Note that the numbers of the contingency tables are translated to a color scale to ease readability.

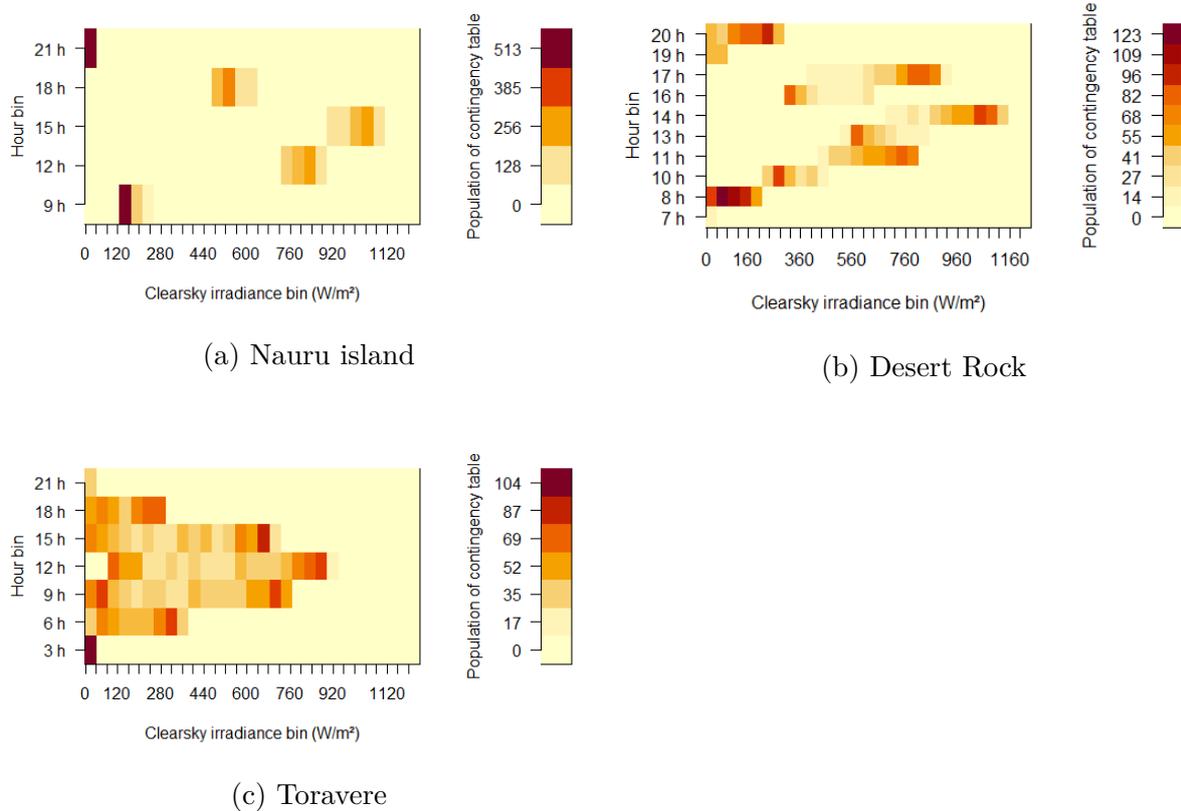
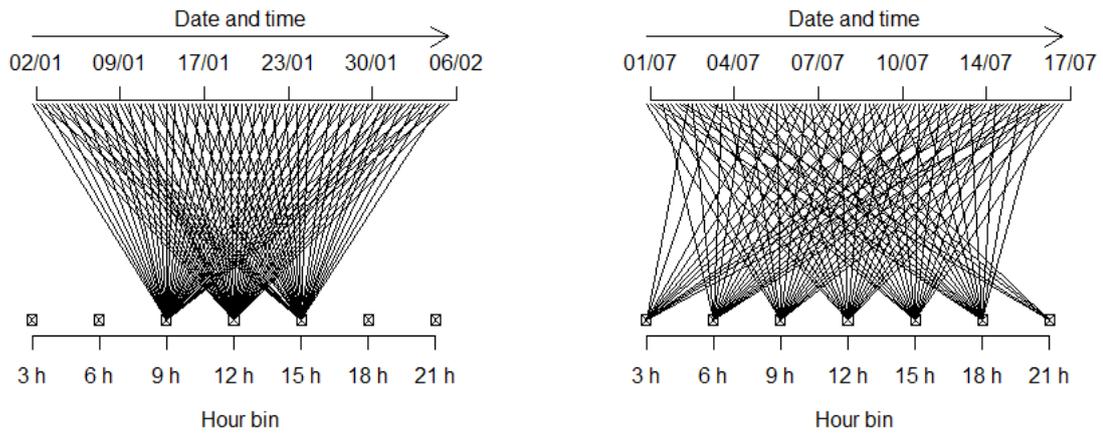


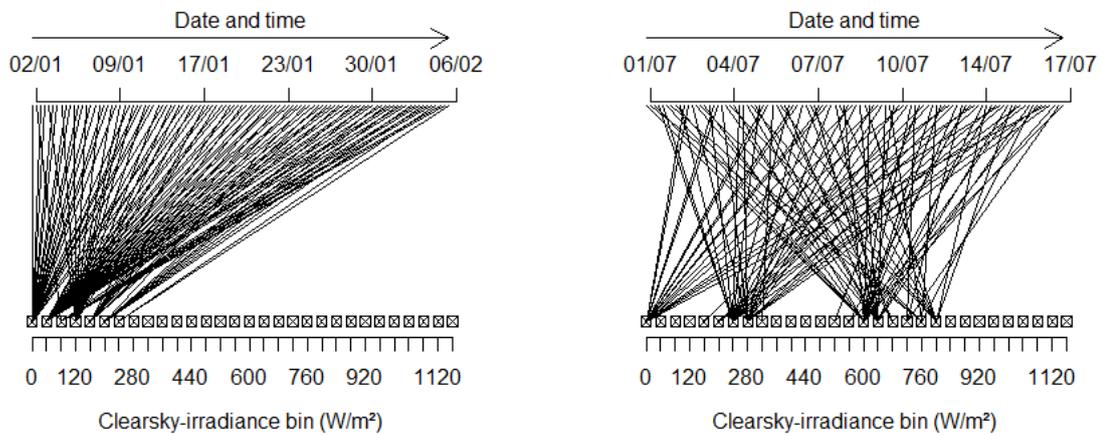
Figure 2.6.3: Contingency tables related to the two binning processes for three sites.

As shown by Figure 2.6.3, the binning process appears very different between the two approaches. Moreover, this difference varies between locations. For instance, in Nauru island, the binning process is more or less equivalent. Indeed, there are no clearsky irradiance bins of CSD-CLIM which corresponds to several hour bins of CH-PeEn. Conversely, this is not the case for in Desert Rock or Toravere. A possible explanation of these discrepancies may come from the latitude of the considered site and more precisely from the seasonal patterns experienced by a site. As an illustration, the sites of Nauru, Desert Rock and Toravere have respectively the following latitudes : -0.52° (Nauru island), 36.62° (Desert Rock) and 58.25° (Toravere). It can be argued that the further the site is from the equator, the more prominent the seasonal effect. It must be stressed here that this seasonal effect is ignored by CH-PeEn model. In other words, the further the site is from the equator, the more different the two binning processes are.

In practice, the difference between the two approaches varies also according to the seasons of the year. An example of this significant difference is illustrated for the site of Toravere in Figure 2.6.4.



(a) CH-PeEn : binning process for Toravere winter days (b) CH-PeEn : binning process for Toravere summer days



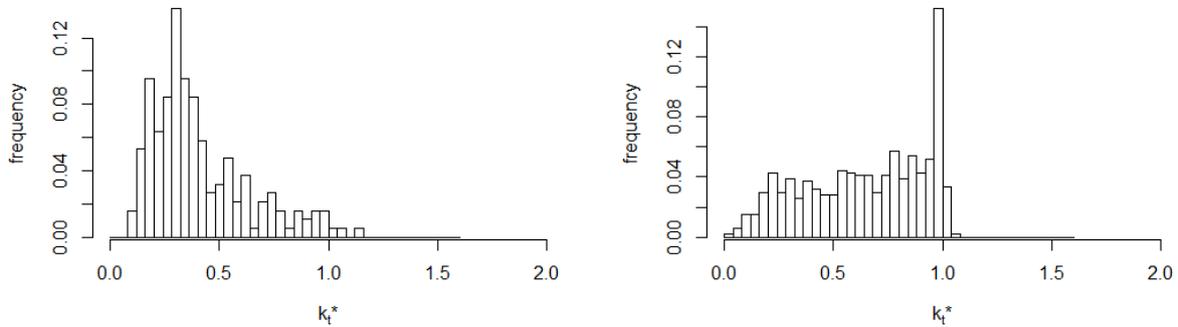
(c) CSD-CLIM : binning process for Toravere winter days (d) CSD-CLIM : binning process for Toravere summer days

Figure 2.6.4: Differences in the binning process between CH-PeEn and CSD-CLIM models, year 2017.

It is obvious from Figure 2.6.4 that the implementation of the binning process is very different between the two models. As shown by Figures 2.6.4a and 2.6.4b, the winter and summer seasons are treated the same way by CH-PeEn, except that more hour bins are involved during summer days. For CSD-CLIM, in winter (see Figure 2.6.4c), all observations are grouped into low clear-sky irradiance bins while they are more equally apportioned in summer (see Figure 2.6.4d).

In fact, CH-PeEn relies on a very strong implicit assumption i.e. for a fixed hour, the distribution of the clear-sky indices should be the same at each time of the year. This assumption does not hold completely true in various cases. Not surprisingly, in locations where the seasonal pattern is strong, the differences in the predictive CDFs (used to compute the $CRPS_n$) related to the winter and summer seasons should be non-negligible. An illustration of such a situation is presented below in Figure 2.6.5. The clear-sky indices of Toravere of a fixed hour (09h-12h UTC) are distributed according to the level of the clear-sky irradiance (low clear-sky irradiances, corresponding approximately to winter months, in Figure 2.6.5a and high clear-sky irradiances, corresponding approximately to

summer months, in Figure 2.6.5b).



(a) Months from October to March
(clear-sky irradiances inferior to 250 W/m^2)

(b) Months from March to October
(clear-sky irradiances superior to 250 W/m^2)

Figure 2.6.5: Clear-sky index distributions for Toravere, averaged on the 3-hour window[09h-12h] UTC.

Figure 2.6.5 shows a significant difference between the two distributions. This difference cannot be taken into account by CH-PeEn since it groups the GHI data according to the hour of the day. Thus, the weak point of CH-PeEn is to aggregate together cases that a clear-sky model could easily discriminate. This also explains the difference in the *CRPS* results between CH-PeEn and CSD-CLIM. The mixture of situations that are statistically different made by CH-PeEn increases its *CRPS*. Indeed, it is known that the local hour is not directly correlated with the seasonal and daily cycles of the sun irradiance. Consequently, the variability of the k_t^* bins used for the CH-PeEn includes the variability due to the solar declination, the time equation or the level of Aerosol Optical Depths (AODs). Conversely, the binning made by CSD-CLIM is finer as it is governed by the specification of the number of bins used by the binning process (see Appendix 2.A). Put differently, CH-PeEn only takes profit from the daily periodicity of the climatology, whereas the results of this study tend to show that other periodicities (like the seasonality of the sun path) are non-negligible.

2.7 Conclusions

In this work, a new climatology reference model is proposed to benchmark solar irradiance probabilistic forecasts. This new model called CSD-CLIM (for Clear-Sky Dependent climatology) is part of the class of climatology benchmark models. CSD-CLIM is evaluated against two existing climatology reference models. The first one is the naive climatology model while the second one is a recommended model in the solar forecasting community namely the CH-PeEn (for Complete History Persistence Ensemble) proposed by Yang (2019). After having defined a set of properties that a benchmark model should have, we have shown that CSD-CLIM, similarly to the naive climatology and CH-PeEn,

- is easy-to-implement,

- has an implementation that depends only on the historical data at hand after a correct selection of a binning parameter,
- has a performance which is time invariant.

Besides, it was also demonstrated that, unlike the naive climatology, CH-PeEn and CSD-CLIM are able to reflect the difficulty of forecasting at a particular location.

More importantly, it was shown that CSD-CLIM achieves the best trade-off between the two most important attributes of a probabilistic forecast namely reliability and resolution. In particular, CSD-CLIM can be qualified as more statistically consistent than CH-PeEn. As such, in terms of overall performance, CSD-CLIM slightly outperforms CH-PeEn. This improved performance is due to a specific binning of the historical irradiance data based on the clear-sky irradiance values.

Finally, and as a conclusion, we can argue that the CSD-CLIM model can be a viable alternative to the CH-PeEn model.

Appendices

2.A Sensitivity Analysis related to the number of bins used by the CSD-CLIM binning process

In this section, an analysis of the sensitivity of *CSD-UNC* related to Nb is conducted. A number $Nb = 30$ has been chosen in this study. However, this choice is arbitrary. It could have a strong impact on the final result and could be questioned. Note that a choice of $Nb = 1$ makes *CSD-UNC* equal to *UNC*. The impact of Nb on *CSD-UNC* is presented in Figure 2.A.1

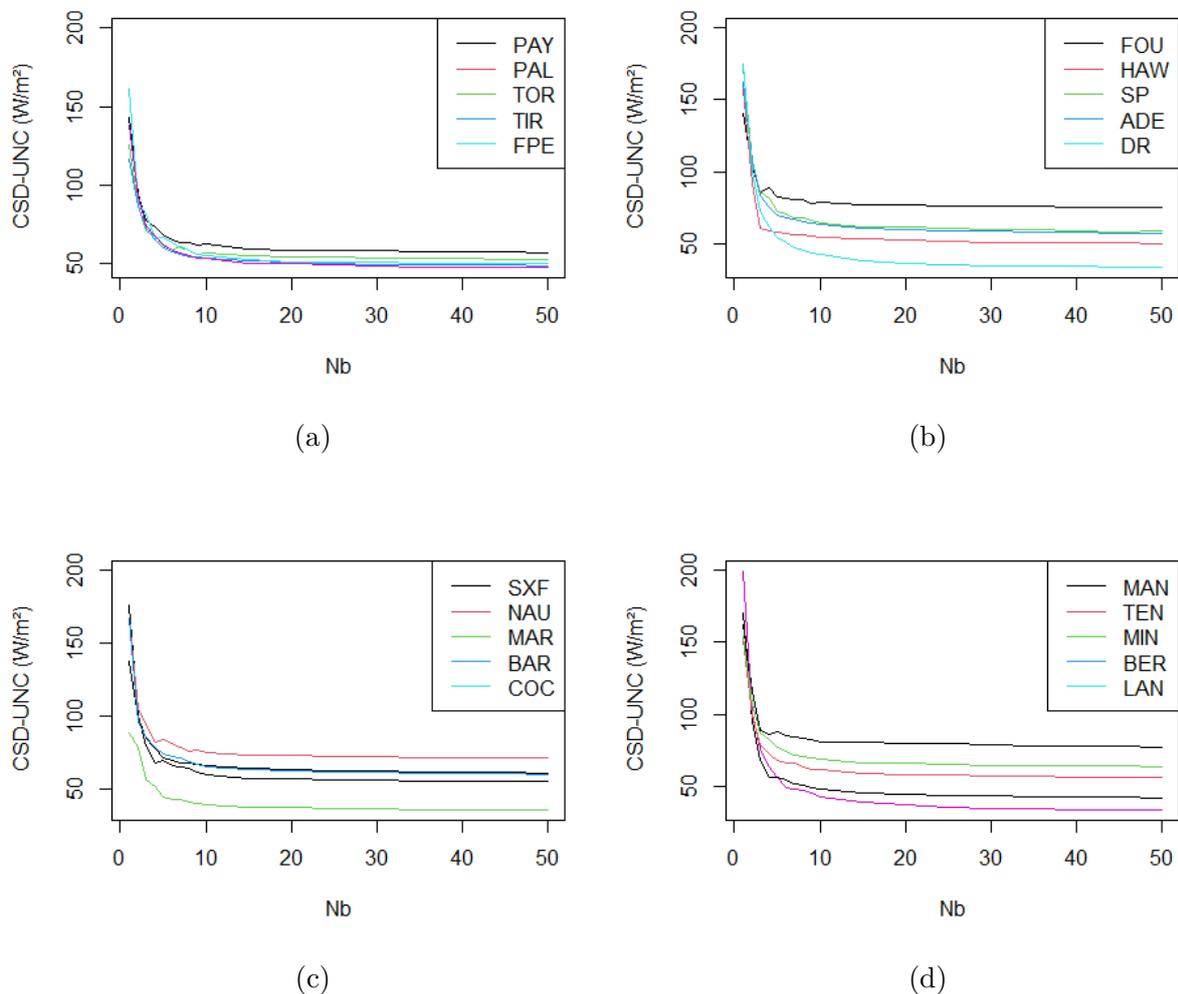


Figure 2.A.1: Impact of Nb on *CSD-UNC*.

Two very different regimes are distinguishable. For all sites, *CSD-UNC* decreases dramatically when Nb varies from 1 to 20. From a number $Nb = 20$, *CSD-UNC* is stable and the choice of Nb is no longer of great importance. Thus, a choice of $Nb > 20$ should be preferred. Our choice of $Nb = 30$ meets this requirement. Note that the number of regime switching (here $Nb = 20$) is not absolute and could depend on the size of the data.

The uncertainty term depends on the variability of the observations and is defined by :

$$UNC_{BS} = \overline{O}(1 - \overline{O}). \quad (2.11)$$

In the case of a continuous variable like GHI, the *Brier score* can be used to evaluate the probability that GHI exceeds a threshold x .

2.B Time-invariance of CSD-CLIM

Since the score *CSD-UNC* of CSD-CLIM is a climatological indicator theoretically based on all historical data, its stability is necessarily achieved when the length of the input data is sufficiently large. Nonetheless, in this section, we investigate the dependency of *CSD-UNC* for different cases of input data.

Four sites of the study i.e. Payerne, Sioux Falls, Tenerife and Bermuda island which experience different sky conditions (see Table 2.3.1) have been selected. The *CSD-UNC* was calculated for 3 different periods of 3 years and for 7 historical datasets with different length (from 1 to 7 years). The periods and the lengths of the different datasets are listed respectively in Table 2.B.1 and Table 2.B.2

Site	Period		
	period 1	period 2	period 3
Payerne	2011-2013	2014-2016	2017-2019
Sioux Falls	2010-2012	2013-2015	2016-2018
Tenerife	2012-2014	2015-2017	2018-2020
Bermuda Islands	2004-2006	2007-2009	2010-2012

Table 2.B.1: Periods used for time stability assessment.

Site	Data length						
	1	2	3	4	5	6	7
Payerne	2011	2011-2012	2011-2013	2011-2014	2011-2015	2011-2016	2011-2017
Sioux Falls	2010	2010-2011	2010-2012	2010-2013	2010-2014	2010-2015	2010-2016
Tenerife	2012	2012-2013	2012-2014	2012-2015	2012-2016	2012-2017	2012-2018
Bermuda Islands	2004	2004-2005	2004-2006	2004-2007	2004-2008	2004-2009	2004-2010

Table 2.B.2: Data lengths used for time stability assessment.

The resulting *CRPS* are given respectively in Table 2.B.3 and Table 2.B.4

Site	<i>CSD-UNC</i> (W/m ²)		
	period 1	period 2	period 3
Payerne	59.3	60.5	57.9
Sioux Falls	60.1	61.0	61.2
Tenerife	35.1	37.6	35.9
Bermuda Islands	59.0	57.8	56.2

Table 2.B.3: Sensitivity of *CSD-UNC* on data period.

Site	<i>CSD-UNC</i> (W/m ²)						
	1	2	3	4	5	6	7
Payerne	55.6	57.6	59.3	59.3	59.3	60.0	59.8
Sioux Falls	60.4	61.3	60.1	61.5	61.6	61.4	60.8
Tenerife	38.0	35.6	35.1	36.4	36.2	36.5	36.7
Bermuda Islands	56.9	58.0	59.0	59.5	59.6	58.6	58.6

Table 2.B.4: Sensitivity of *CSD-UNC* on data length.

As shown by these tables, *CSD-UNC* is not strongly dependent on the chosen period or length of the dataset used to calculate it. In this work, the data granularity was 3h. However, it must be stressed that the stability in the *CRPS* results can be improved provided that the time resolution of the data increases. We recall that in a practical case, if the dependency on the input data is found strong, the recommendation should be to extend the length of the input data in order to get closer to the climatological mean.

2.C Reliability assessment using PIT Histograms

In this section, the reliabilities of the three considered models are assessed with the help of PIT histograms. The PIT is the value of the predictive CDF for a particular observation O i.e. $\hat{\mathcal{F}}(O)$. One can plot the histogram of the PIT values computed over a testing set of observations. A perfectly reliable forecast is characterized by a flat PIT histogram with a given theoretical frequency. The consistency bars recommended by Bröcker and Smith (2007b) are included to see how much deviation from the ideal (uniform) theoretical line can be seen as acceptable, in view of the testing sample size. The interested reader is referred to Gneiting et al. (2005) for an exhaustive presentation of this diagnostic tool.

For the sites of Payerne (“PAY”), Sioux Falls (“SXF”), Tenerife (“TEN”) and Bermuda Islands (“BER”), the PIT histograms produced by the three models on the test data are compared in Figure 2.C.1.

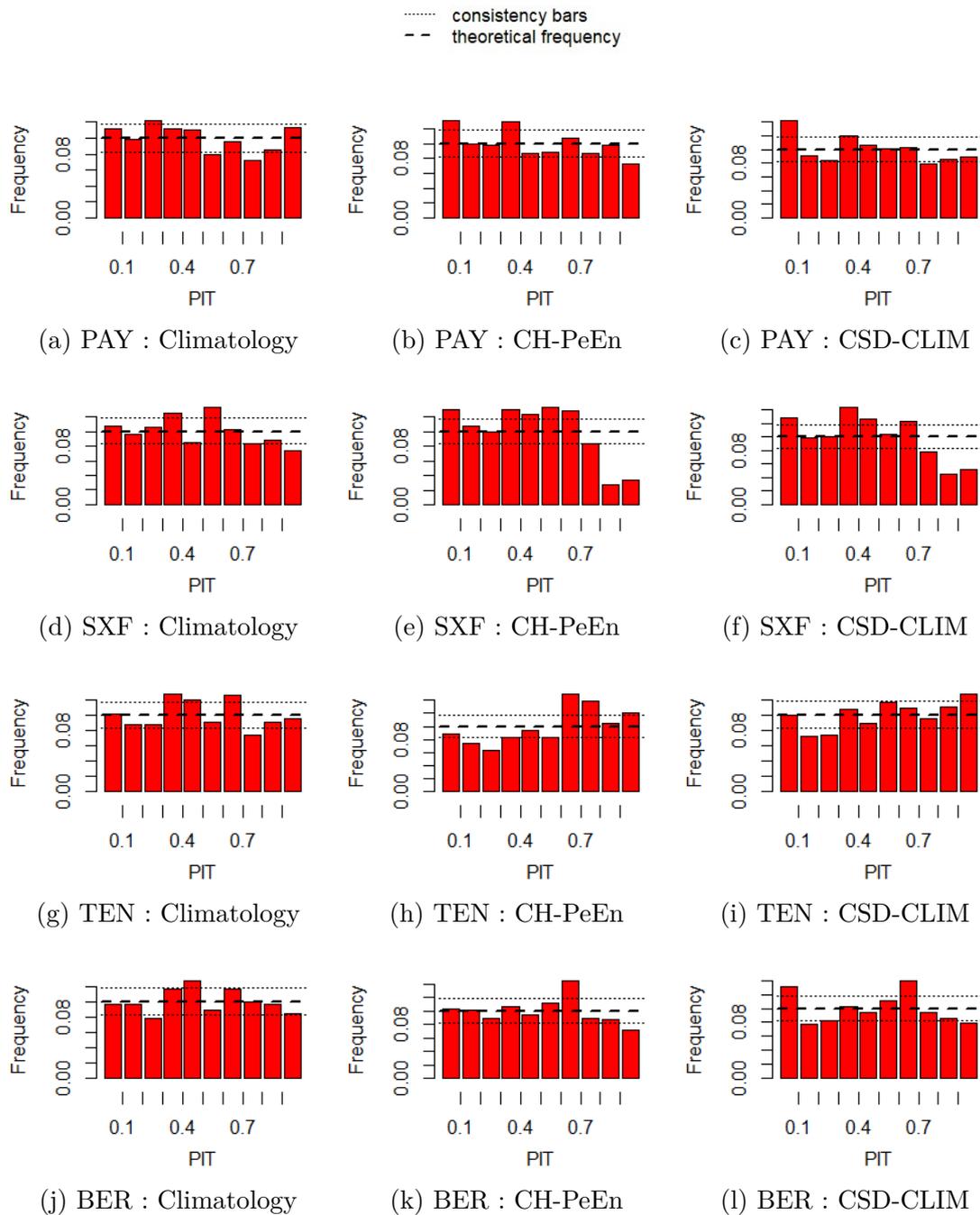


Figure 2.C.1: PIT Histograms of the three benchmark models evaluated on the test data.

As shown by Figure 2.C.1, the visual inspection of the PIT histograms evaluated on the test data shows that, irrespective of the site, and given the testing sample size, the naive climatology model seems to be the more reliable. Conversely, more important deviations from the ideal theoretical line are observed in the case of the two other models albeit it appears that these deviations are slightly more pronounced for CH-PeEn. Although one must be careful when drawing conclusions from this visual diagnostic tool, we can reasonably state that this reliability assessment with the help of PIT histograms leads to the same results issued from the decomposition of the *CRPS* (see section 2.5.3).

Chapitre 3

Gains effectifs des approches probabiliste et ensembliste

Résumé

Dans ce chapitre, on cherche à mesurer les gains effectifs des approches probabiliste et ensembliste sur un cas d'étude pratique. Ainsi, ce chapitre fait écho au chapitre 1, qui avait pour but de quantifier les gains potentiels de qualité. En accord avec les conclusions de ce dernier, le *CRPS* est utilisé pour mesurer la qualité des prévisions. Afin de mesurer les gains probabilistes effectifs, ce chapitre propose l'utilisation d'une méthodologie, appelée MC-CRPS, ayant pour but de mettre en évidence les gains en qualité attribuables aux différents moments statistiques d'une prévision. Ce chapitre reprend très largement l'article nommé "Added-value of Ensemble Prediction System on the quality of solar irradiance probabilistic forecasts" publié dans la revue internationale à comité de lecture "Renewable Energy" (Le Gal La Salle et al., 2020). Les seules modifications apportées sont l'ajout de compléments permettant de faciliter l'intégration de ce chapitre dans le travail de thèse (section 3.8), ainsi que la modification de certaines notations afin de les rendre cohérentes avec l'ensemble du travail.

Sommaire

3.1	Introduction	87
3.2	Building probabilistic forecasts	90
3.2.1	Statistical techniques used to generate probabilistic forecasts	91
3.2.2	Obtaining probabilistic forecasts from deterministic forecasts (Deterministic-based approach)	95
3.2.3	Obtaining probabilistic forecasts from ensemble forecasts (Ensemble-based approach)	95
3.3	Verification of the probabilistic forecasts	96
3.3.1	Attributes for a skillful probabilistic model	96
3.3.2	The <i>CRPS</i>	97
3.3.3	Contributions of the statistical moments of the forecast distribution to the <i>CRPS</i>	98
3.4	Case studies	100
3.4.1	Measurements	100

3.4.2	Forecasts	101
3.5	Results	101
3.5.1	Overall performance of the methods	102
3.5.2	Detailed insight through the decomposition of the <i>CRPS</i>	103
3.5.3	Detailed insight through the CRPS Moments-Contributions	105
3.6	Discussion	106
3.6.1	Deterministic-based approach versus ensemble-based approach	106
3.6.2	Discussion related to CRPS Moments-Contributions	107
3.7	Conclusions	112
	Appendices	113
3.A	Data quality check	113
3.B	Bias and standard deviation of EPS members distribution and observations for the six sites	116
3.C	Selection of the optimal α	118
3.8	Compléments	120
3.8.1	<i>CRPS skill scores</i> par rapport au modèle CSD-CLIM	120
3.8.2	Gains effectifs de qualité obtenus grâce à l’approche probabiliste	121

Added-value of Ensemble Prediction System on the quality of solar irradiance probabilistic forecasts

Abstract

Accurate solar forecasts is one of the most effective solution to enhance grid operations. As the solar resource is intrinsically uncertain, a growing interest for solar probabilistic forecasts is observed in the solar research community. In this work, we compare two approaches for the generation of day-ahead solar irradiance probabilistic forecasts. The first class of models termed as deterministic-based models generates probabilistic forecasts from a deterministic value of the irradiance predicted by a Numerical Weather Prediction (NWP) model. The second type of models denoted by ensemble-based models issues probabilistic forecasts through the calibration of an Ensemble Prediction System (EPS) or from information (such as mean and variance) derived from the ensemble. The verification of the probabilistic forecasts is made using a sound framework. A numerical score, the Continuous Ranked Probability Score (*CRPS*), is used to assess the overall performance of the different models. The decomposition of the *CRPS* into reliability and resolution provides a further detailed insight into the quality of the probabilistic forecasts. In addition, a new diagnostic tool which evaluates the contribution of the statistical moments of the forecast distributions to the *CRPS* is proposed. This tool denoted by MC-*CRPS* allows identifying the characteristics of an ensemble that have an impact on the quality of the probabilistic forecasts. The assessment of the different models is done on several sites experiencing very different climatic conditions. Results show a general superior performance of ensemble-based models as the gain in forecast quality measured by the *CRPS* compared to deterministic-based models ranges from 4% to 16% depending on the site.

3.1 Introduction

Operations of electrical power systems are becoming more challenging as the share of solar energy increases. In particular, due to the intrinsic variability of the solar resource, high penetration of solar power generation into the electrical grid may put in danger the grid supply-demand balance. Energy storage systems (EES) are one of the means used to ensure the grid stability. Notwithstanding, accurate PV power forecasting is a cost-effective way to size and operate ESS optimally. Consequently, PV power forecasts facilitate the large-scale integration of solar energy into the grid. In addition, for energy trading, accurate PV power forecasts are also required because penalties in proportion with the forecast errors are applied.

In this study, however, we focus on the global horizontal solar irradiance (GHI) forecasts instead of PV power forecasts. The present work constitutes thus a first step in assessing the contribution of the proposed methodologies for improving the quality of the PV power forecasts and of their potential gain for improved grid operations. Day-ahead GHI forecasts are treated here as they have been considered essential to secure the power grid (Pierro et al., 2019). Moreover, we propose to work on probabilistic forecasting in order to estimate the uncertainty associated to day-ahead GHI forecasts. This additional knowledge permits for instance grid operators to improve their decisions regarding the

grid operations. The interested reader can refer to Zhu et al. (2002) or Buizza (2008) to understand the benefits of a probabilistic forecast against a deterministic one.

Day-ahead GHI forecasts are classically generated by Numerical Weather Predictions models (NWP). For instance, The Integrated Forecasting System (IFS) model of the European Centre of Medium-Range Weather Forecasts (ECMWF) provides day-ahead GHI forecasts Leutbecher and Palmer (2008). The forecasts can take either the form of a deterministic forecast or an ensemble forecast denoted by the term Ensemble Prediction System (EPS). EPS consists in a set of several perturbed forecasts of irradiance, each representing a possible future state of the atmosphere. If an EPS gives an important information about the uncertainty associated to a forecast, it requires a high computational cost. Thus, the added value of EPS for probabilistic forecasting needs to be determined to justify their computation.

We propose below to conduct a bibliographic survey related to day-ahead solar forecasts with a special emphasis on the use of NWP outputs to generate probabilistic forecasts. One of the first approach used to generate day-ahead probabilistic irradiance forecasts was proposed by Lorenz et al. (2009b). In this work, a Gaussian distribution of the error of the ECMWF-IFS deterministic irradiance forecast was used to generate prediction intervals. Alessandrini et al. (2015) developed an analog statistical method approach applied to a set of explanatory weather variables (GHI, cloud cover, air temperature, etc.) provided by the NWP Regional Atmospheric Modeling System (RAMS) to generate probabilistic PV power forecasts for three solar farms located in Italy. Zamo et al. (2014b) proposed two statistical approaches to generating probabilistic forecasts of daily PV production from information provided by Météo France's EPS, PEARP. The first approach makes use of the PEARP control member as unique input to quantile regression methods while the second one averages the set of quantiles calculated from each of the 35 members of the PEARP ensemble. Bacher et al. (2009) used a weighted quantile regression (WQR) technique to compute up to 24h ahead probabilistic PV forecasts. In addition to lagged PV measurements, the WQR model used also a NWP-based GHI deterministic forecast. Lauret et al. (2017) used the IFS model to produce quantile forecasts of solar irradiance and Iversen et al. (2014) introduces the idea of modeling uncertainty by stochastic differential equations from a NWP-based deterministic forecast provided by the Danish Meteorological Institute. Bakker et al. (2019) proposed a comparison of seven statistical regression models to issue GHI probabilistic forecasts from the deterministic numerical weather prediction (NWP) model HARMONIE-AROME (HA) and the atmospheric composition model CAMS.

It must be noted that the above cited works make use of deterministic information extracted from NWP models to generate probabilistic forecasts with the help of statistical techniques like quantile regression or analog ensemble. Others authors like Sperati et al. (2016) proceeded differently. In their work, Sperati et al. (2016) generated up to 72h probabilistic forecasts from the raw EPS provided by the ECMWF. In this study, two post-processing methods (also called calibration techniques) applied to the initial raw ensemble were used to further improve the quality of the probabilistic forecasts. Massidda and Marrocu (2018) went a little bit further and proposed a methodology to combine ECMWF ensemble and the high-resolution IFS deterministic forecast.

If we extend our bibliographic survey to the probabilistic predictions of other weather variables such as wind, temperature or precipitation, more publications can be found on how to use information from NWP models to generate probabilistic forecasts. For

example, Pinson (2012) and Pinson and Madsen (2009) suggested a framework for the calibration of wind ensemble forecasts. Junk et al. (2015) proposed an original calibration model for wind-speed forecasting applied to ECMWF-EPS based on the combination between Nonhomogeneous Gaussian Regression and Analog Ensemble Models. Likewise, Hamill and Whitaker (2006) suggested an adaptation of the analog ensemble technique for the calibration of ensemble precipitation forecast, using the statistical moments of the distribution such as mean and spread of the members as predictors.

Wilks (2006), followed in his methodology by Williams et al. (2014), compared several post-processing techniques of weather EPS forecasts, such as ensemble dressing, Logistic Regression, Nonhomogeneous Gaussian Regression (“NGR”) and Rank-Histogram recalibration. The reader can refer to Gneiting et al. (2005) and Lerch and Thorarinsdottir (2013) for more details regarding the parametric calibration of ensemble forecasts with techniques like NGR with a special emphasis on the choice of the type of the parametric distribution used by the regression technique. Finally, the interested reader should consult the reference book (Vannitsem et al., 2018b), which proposes a summary of the common probabilistic forecasting ensemble-based models with their respective pros and cons.

Based on this bibliographic survey, two different approaches for day-ahead GHI probabilistic forecasting with the help of NWP models can be identified, which we denoted here by approaches 1 and 2 :

1. Approach 1 referred herein as *deterministic-based models* : the probabilistic forecast is computed from deterministic NWP predictors with the help of statistical methods. Linear Quantile Regression and Analog Ensemble techniques are particularly attractive to implement this methodology,
2. Approach 2 referred herein as *ensemble-based models* : the estimation of the forecast is made through the calibration of an EPS or from information (for example mean or spread) inferred from the ensemble. For instance, calibration techniques like Nonhomogeneous Regression can be used to improve the raw ensemble EPS. Also, methods based on Linear Quantile Regression and Analog methods can be used to produce probabilistic forecasts from the mean and spread of the ensemble.

It must be stressed however that, to the best of our knowledge, no previous works have been dedicated to the comparison of the two approaches and particularly in the realm of solar probabilistic forecasts. In this work, our main goal is therefore to assess the relative merits of each approach for day-ahead GHI probabilistic forecasts. Besides, we would like to highlight the possible added-value brought by EPS for probabilistic forecasting. Indeed, it is well known that the generation of such ensemble necessitates high computing capacities compared to a single deterministic forecast that is fed into a statistical method to produce the probabilistic forecasts. More precisely, it should be noted that the calculation cost is not the same to produce only the control member of EPS or the whole set of members.

To understand the benefits associated with the usage of EPS, we propose in this paper a sound and consistent methodology to evaluate the respective contribution of each approach. First, the quality appraisal of the different models will be made according the verification framework proposed by Lauret et al. (2019). This framework (which is not consistently proposed in the literature) is based on visual diagnostic tools and numerical scores like the Continuous ranked Probability Score (*CRPS*) which permits to objectively

rank the competing forecasting methods. However, this classical verification framework is not sufficient to completely explain the contribution of the statistical moments of the forecast distributions to the forecast quality. That is why we propose in a second step a new tool that evaluates the accuracy of all moments of the forecast distribution and its contribution to the *CRPS*. We hope that this new diagnostic tool will provide a more in-depth understanding of the performance of each approach. To this end, we evaluate models that generate day-ahead GHI probabilistic forecasts on 6 sites that experience different sky conditions. The probabilistic models are built :

1. With only the control member of the EPS as a deterministic predictor (deterministic-based approach),
2. With a deterministic predictor inferred from the whole set of EPS's members. The first statistical moment (mean of the members) can be such a deterministic predictor (ensemble-based approach),
3. With several predictors inferred from the ensemble like the mean and the variance of the ensemble (ensemble-based approach).

We propose the following structure for the paper. Section 3.2 introduces the different forecasting models while section 3.3 briefly presents the diagnostic tools used for the verification of probabilistic forecasts. Section 3.4 presents the case studies and details the data used to evaluate the different probabilistic models. Section 3.5 provides a detailed assessment of the performance of the different methods. Finally, a discussion will be conducted in section 3.6, trying to understand the pros and cons of each forecasting methods and the factors impacting the forecast quality.

3.2 Building probabilistic forecasts

Regarding probabilistic forecasts of continuous predictand like GHI, a probability statement i.e. either a Probability Distribution Function (PDF) \hat{f} or a Cumulative Distribution Function (CDF) \hat{F} encodes the uncertainty of the forecast. In this work, three ways to estimate this CDF or PDF are considered: parametric PDFs, discrete quantile estimates of a CDF via a non-parametric method and CDF derived from EPS.

In this study, the EPS is provided by the European Centre of Medium-Range Weather Forecasts (ECMWF). It corresponds to 50 perturbed members and a control run (unperturbed member) (Leutbecher and Palmer, 2008) that give the cumul of the GHI with a 3 hours time step. This leads to a total of $M = 51$ members. An EPS can be seen as discrete estimates of a CDF when they are sorted in ascending order. Lauret et al. (2019) discussed three ways to associate these sorted members to cumulative probabilities. In this work, we chose the uniform distribution which consists in a uniform spacing of the members and a linear interpolation between the members. More precisely, this choice assigns a probability mass of $1/(M + 1)$ between two members and for events that fall outside of the ensemble. Using this definition, the i^{th} ensemble member can be interpreted as a quantile forecast with a probability level equal to $\tau = \frac{i}{M+1}$. Put differently, the ECMWF ensemble forecasts are in the form of 51 equally spaced quantiles with probability levels $\tau = \frac{1}{52}, \frac{2}{52}, \dots, \frac{51}{52}$. This construction is illustrated in Figure 3.2.1, for an EPS with 4 members. In the following, we present first the different statistical techniques

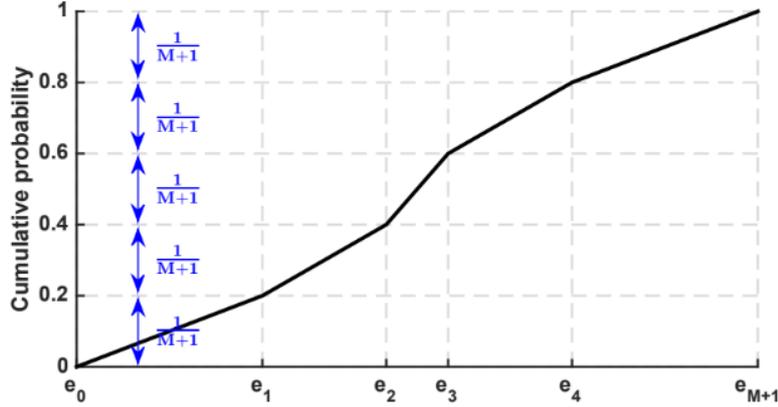


Figure 3.2.1: Illustration of an uniform construction of a CDF from an ensemble of $M = 4$ members (e_1, e_2, e_3, e_4). The tails of the CDF are bounded by e_0 and e_{M+1} which correspond to the minimum and the maximum of the climatology.

used to estimate the uncertainty of the forecasts. Secondly, we detail the two approaches introduced in section 3.1.

3.2.1 Statistical techniques used to generate probabilistic forecasts

The linear quantile regression (LQR) technique

This method estimates the quantiles of the cumulative distribution function $\hat{\mathcal{F}}$ of some response variable X (also called predictand) by assuming a linear relationship between the quantiles of X , namely Q_τ and a set of explanatory variables Y (called predictors):

$$Q_\tau = \beta_\tau Y + \varepsilon, \tag{3.1}$$

where β_τ is a vector of parameters to optimize for each probability level τ and ε represents a random error term.

Following Koenker and D'Orey (1994), the vector $\hat{\beta}_\tau$ that defines each quantile is obtained as the solution of the following minimization problem:

$$\hat{\beta}_\tau = \arg \min_{\beta} \sum_{n=1}^{N_A} \rho_\tau (O_n - \beta Y_n). \tag{3.2}$$

where N_A is the number of pairs of observed predictand O_n , set of predictors Y_n taken from the training set. $\rho_\tau(u)$ is the quantile loss function defined as :

$$\rho_\tau(u) = \begin{cases} u\tau & \text{if } u \geq 0, \\ u(\tau - 1) & \text{if } u < 0, \end{cases} \tag{3.3}$$

with τ representing the quantile probability level. Hence, in quantile regression, the quantiles are estimated by applying asymmetric weights to the mean absolute error.

Thus, the quantity $\hat{F}_\tau = \hat{\beta}_\tau Y$ is the estimation of the τ^{th} quantile obtained by the LQR method.

It must be noted that the quantile regression method estimates each quantile separately (i.e. the minimization of the quantile loss function is made for each τ separately). As a consequence, one can obtain quantile regression curves that may intersect, i.e. $\hat{F}_{\tau_1} > \hat{F}_{\tau_2}$ when $\tau_1 < \tau_2$. To avoid this issue during the model fitting, we used the rearrangement method described by Chernozhukov et al. (2010).

Figure 3.2.2 shows some quantiles estimates of the CDF of the predictand X (here GHI) as a function of the day-ahead forecasted GHI. Hence, in this case, the predictor Y is the predicted irradiance which will be represented in this work either by the ECMWF control member or the mean of the ECMWF ensemble (see Table 3.2.2 below). This example shows that the forecast uncertainty depends on the level of the predicted irradiance. More precisely, and as shown by Figure 3.2.2, the dispersion of points is lower for values of predicted irradiance close to 0 W/m^2 and greater for values between 40 and 100 W/m^2 .

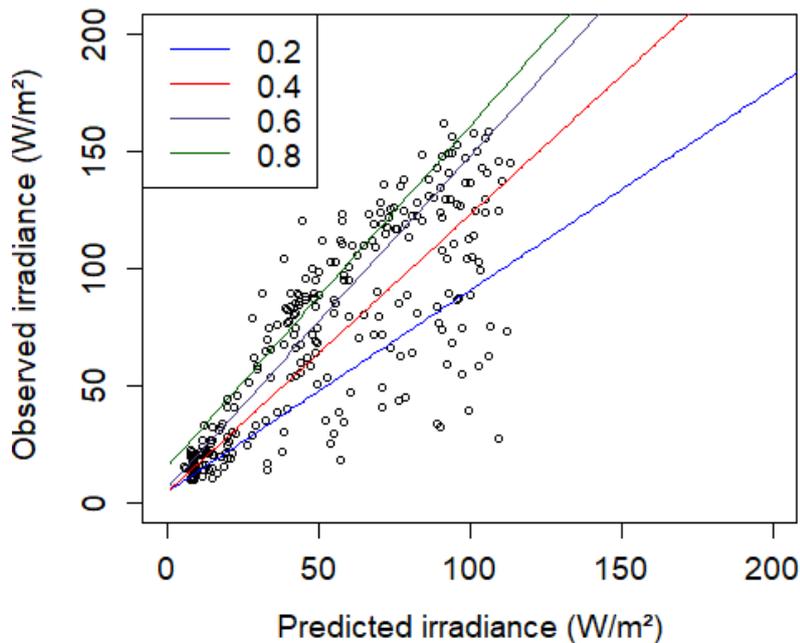


Figure 3.2.2: Observed GHI vs. the predicted day-ahead GHI. The lines are the estimates of the quantiles with probability levels of 0.2, 0.4, 0.6 and 0.8. Data are from the training period of Hawaii. Observed and predicted GHI are averaged on the 3-hour window [17h-20h] local time.

The Analog Ensemble (AnEn) technique

The analog ensemble technique is now quite a standard in the energy meteorology forecasting community (Delle Monache et al., 2013; Hamill and Whitaker, 2006). Similarly to the LQR method, the Analog Ensemble technique is a non-parametric method that can be used to estimate the predictive CDF of the predictand.

Considering a training set of N_A ordered (sorted by forecasts) pairs of GHI observa-

tions/GHI forecasts $(O_n, \hat{F}_{d,n})_{n=1, \dots, N_A}$, the procedure for determining the forecast CDF is as follows:

1. For a new forecast taken from a testing set, calculate its distance from every past forecast order and find the rank R of the past forecast that is the closest to the new forecast,
2. Form an ensemble by selecting the $2\alpha + 1$ past training observations O_k having their ranks k inside the interval $[R - \alpha, R + \alpha]$,
3. Compute the predictive CDF at a specific value x of the predictand using the following equation:

$$\hat{F}(x) = P(X \leq x) = \frac{1}{2\alpha + 1} \sum_{k=1}^{2\alpha+1} H(x - O_k), \quad (3.4)$$

where X is the random value related to the predictand (here GHI) and H is the Heaviside or step function. The effectiveness of the method is strongly dependent on the value of α . It is proposed here to take $\alpha = 0.02N$. This choice has been motivated by a preliminary study made on the training period. Appendix 3.C details the selection of the optimal value of α . Finally, as for the linear quantile regression, note that the GHI forecasts used in the AnEn technique will be given either by the ECMWF control member or the mean of the ensemble (see Table 3.2.2 below).

The Nonhomogeneous truncated Gaussian Regression technique (t_NGR)

The NGR technique also called in some studies ‘‘Ensemble Model Output Statistics’’ (EMOS) has been introduced by Gneiting et al. (2005) for probabilistic forecasting of weather variables. This technique is dedicated to the post-processing of ensemble forecasts produced by an EPS. The NGR technique builds the predictive PDF of the predictand X from a normal PDF. As such, this kind of model can be termed as a parametric model. The predictive pdf \hat{f} estimated by the NGR method is given by:

$$\hat{f} \sim \mathcal{N}\left(a + \sum_{i=1}^M (b_k E_i), c + dS^2\right), \quad (3.5)$$

where M is the number of members, m_k is the k^{th} member and σ^2 is the variance of the ensemble members distribution. The free parameters a, b_1, \dots, b_M, c and d are determined with the help of an optimization procedure. In this work, and following Gneiting et al. (2005), these parameters are calculated by minimizing the *CRPS* over a training period (see section 3.3.2 for details regarding *CRPS*). Furthermore, as GHI is a necessarily positive quantity, we propose, in this work, a variant of the NGR technique namely a truncated version (at 0) of the nonhomogeneous gaussian regression. In the following, the corresponding model is denoted as t_NGR.

The Nonhomogeneous Regression of Generalized Extreme Value technique (NR GEV)

One can question the choice of a Gaussian distribution in the t_NGR technique. Indeed, the distributions of observations for a fixed forecasting level are actually non-Gaussian. Two examples for the studied sites are presented in Figure 3.2.3.

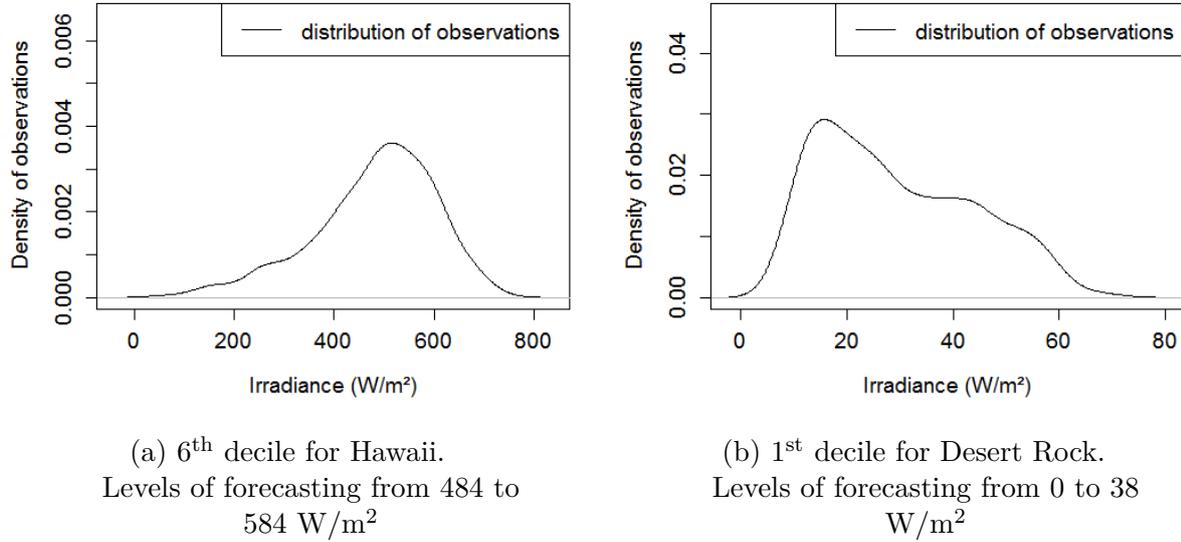


Figure 3.2.3: Example of distributions of observations for a fixed level of forecasting.

On these specific examples, the distributions of observations are clearly non-Gaussian and the consideration of other types of distributions may improve the skills of the forecast.

As pointed out in Lerch and Thorarinsdottir (2013) and Baran and Lerch (2018), other types of parametric distributions can be used to deal with this issue. Here, a Non homogeneous Regression approach with Generalized Extreme Value distributions is proposed to estimate the PDF of the predictand X . The PDF of a generalized Extreme value distribution for a specific value x of the predictand GHI is defined as :

$$\hat{f}(x) = \begin{cases} \frac{1}{\eta} \left[1 + \xi \left(\frac{x-\mu}{\eta} \right) \right]^{(-\frac{1}{\xi})-1} \exp \left(- \left[1 + \xi \left(\frac{x-\mu}{\eta} \right) \right]^{-\frac{1}{\xi}} \right) & \xi \neq 0, \\ \frac{1}{\eta} \exp \left(-\frac{x-\mu}{\eta} \right) \exp \left[- \exp \left(-\frac{x-\mu}{\eta} \right) \right] & \xi = 0. \end{cases} \quad (3.6)$$

The parameters μ , η and ξ are to be determined by optimizing the CRPS over the training period. We followed the framework of Scheuerer (2014) and Yuen et al. (2018) to set these coefficients. Following this procedure, the mean μ and the scale parameter η of the final distributions are determined by linear regression, and depends only on variables inferred from the EPS. The mean is a linear combination of the mean of the members and the fraction of members which predict exactly zero. The scale parameter η depends on the “Gini’s mean difference” (a measure of the variability closely related to the spread of the members, see Yitzhaki (2003) for details). Note that the shape parameter is taken as a constant. Thus, the minimization of the *CRPS* yields the linear coefficients for the mean μ and the scale parameter η as well as the value of the shape parameter ξ . Note that the two techniques namely t_NGR discussed above and NR GEV discussed here are part of a family of parametric methods named Nonhomogeneous Regression (“NR”).

3.2.2 Obtaining probabilistic forecasts from deterministic forecasts (Deterministic-based approach)

Some of the techniques presented in section 3.2.1, namely the Linear Quantile Regression (LQR) and the Analog Ensemble (AnEn) techniques, are capable of generating a probabilistic forecast from a deterministic predictor.

In our study, and regarding the deterministic-based approach, the control member of ECMWF-EPS is the predictor variable Y of the LQR technique and it will be the forecast used in the AnEn procedure. The corresponding probabilistic models are denoted respectively as LQR_c and $AnEn_c$ in the following.

3.2.3 Obtaining probabilistic forecasts from ensemble forecasts (Ensemble-based approach)

From the raw output of ECMWF-EPS

Given a raw ensemble forecast of M members $\{E_i\}_{i=1,\dots,M}$, it seems natural to define directly a forecast CDF from this EPS as illustrated in Figure 3.2.1. Note that this definition corresponds to the “uniform” definition of a CDF derived from an ensemble discussed in Lauret et al. (2019).

From information extracted from an EPS

An EPS differs from a deterministic forecast by the multiplicity of predictors. In this work, we propose to assess the quality of two variants of probabilistic models built with information extracted from an EPS.

The first variant will make use of the mean of the ensemble members of the EPS. The use of the mean of members as a deterministic predictor is justified by Table 3.2.1. For all the considered sites depicted in Table 3.4.1, Table 3.2.1 lists the Root Mean Square Error ($RMSE$)¹ of different deterministic predictors extracted from an EPS.

Site	HAW	DR	SP	PAL	TIR	LAN
Any perturbed member	138	75.3	126.5	97.7	110.7	98.8
Control member	135	72.8	91.9	102.9	100.8	93.2
Mean of the members	129.7	67.9	113.8	81.8	92.6	84.3
Median of the members	133.9	69.4	115.5	84.1	94.7	85.6

Table 3.2.1: $RMSE$ (W/m^2) of 4 deterministic forecasts that can be inferred from an EPS: any of the 50 perturbed members of ECMWF ensemble forecast, the control member (unperturbed), the mean of the members and the median of the members. See Table 3.4.1 for the signification of the acronyms of the different sites.

As shown by Table 3.2.1, the mean of all the members turns out to be the best predictor for deterministic forecasting. Hence, to quantify the improvement brought by the first moment estimation (i.e. the mean), two models denoted by LQR_m and $AnEn_m$ based respectively on the LQR and AnEn techniques will be evaluated.

¹ $RMSE$ is a common metric used to assess the accuracy of deterministic forecasts (Hoff et al., 2013)

The second variant will include, in addition to the mean of the members, the spread (i.e. the variance) of the members of the EPS. The `t_NGR` and the NR GEV models described in section 3.2.1 use the first and second moment of the EPS distribution to build the predictive distributions. Furthermore, we also propose to use the LQR technique with a vector Y of predictors given by

$$Y = [\mu, \sigma^2], \quad (3.7)$$

where μ represents the mean of members and S^2 the variance of the ensemble. This method will be referred in this study as `LQRs`. Finally, Table 3.2.2 summarizes the different probabilistic models that will be evaluated in this study.

Approach	Deterministic-based		Ensemble-based			
Predictors	Control member		Mean of members		Mean and spread of members	
Technique	AnEn	LQR	AnEn	LQR	LQR	NR
Model Abbreviation	AnEn _c	LQR _c	AnEn _m	LQR _m	LQR _s	t_NGR NR GEV

Table 3.2.2: Summary of all considered forecasting models and their associated colours with AnEn: Analog Ensemble, LQR: Linear Quantile Regression, NR: Nonhomogeneous Regression. Grey : model using the deterministic-based approach, cyan : model using the ensemble-based approach and the mean of the members as predictor, green : model using the ensemble-based approach and the mean and the spread of the members as predictors.

3.3 Verification of the probabilistic forecasts

In this section, we detail some of the verification tools proposed by Lauret et al. (2019) that will be applied to assess the quality of GHI probabilistic forecasts. Following this work, we will rely on a quantitative score namely the continuous ranked probability score (*CRPS*) and its related skill score, the *CRPSS*, to rank objectively the different methods. Moreover, and based on the recommendations of Lauret et al. (2019), we will provide the decomposition of the *CRPS* into the main attributes that affect the quality of the forecasts. In addition to this decomposition, it is worth noting that we will propose in this work a new way to have detailed insight into the performance of the methods. This new methodology is based on the contribution of the moments (mean, variance, etc.) of the forecast distribution to the *CRPS* (see section 3.3.3 below).

3.3.1 Attributes for a skillful probabilistic model

We recall here briefly the two main attributes that characterize the quality of the probabilistic models namely reliability and resolution (Wilks, 2014; Jolliffe and Stephenson, 2003). Reliability or calibration evaluates the statistical consistency between the forecasts and the observations. In the case of a continuous variable like GHI, a high reliability is obtained if predictive distributions and distributions of observations agree. Resolution refers to the ability of the probabilistic model to discriminate among different forecast situations. More precisely, the more distinct the observed frequency distributions for various forecast situations are from the full climatological distribution, the more resolution the forecast model has. A high quality probabilistic model should issue reliable forecasts

with high resolution. In other words, high reliability is a necessary but not a sufficient condition for a high quality probabilistic forecast. The forecast should also exhibit high resolution. For instance, climatological forecasts are perfectly reliable but exhibit no resolution.

3.3.2 The *CRPS*

In the verification framework proposed by Lauret et al. (2019), the authors recommend the computation of a score like the Continuous Ranked Probability Score (*CRPS*) to evaluate the overall quality of the probabilistic models. We recall here the definition of the *CRPS*.

Definition

The *CRPS* measures the difference between the predicted and observed cumulative distributions functions (CDF) (Hersbach, 2000). The *CRPS* for one forecast and one corresponding observation indexed by n reads as

$$CRPS_n = \int_{-\infty}^{+\infty} \left[\hat{\mathcal{F}}_n(x) - \mathcal{O}_n(x) \right]^2 dx, \quad (3.8)$$

where $\hat{\mathcal{F}}_n(x)$ is the predictive CDF of the predictand X (here the GHI) and $\mathcal{O}_n(x)$ is a cumulative-probability step function that jumps from 0 to 1 at the point where the value of x equals the observation O_n (i.e. $\mathcal{O}_n(x) = 1_{\{x \geq O_n\}}$). The *CRPS* rewards concentration of probability around the step function located at the observed value (Wilks, 2014). In other words, the *CRPS* penalizes lack of resolution of the predictive distributions as well as biased forecasts. The squared difference between the two CDFs is averaged over the N forecast/observation pairs to define the average *CRPS*.

$$CRPS = \frac{1}{N} \sum_{n=1}^N CRPS_n. \quad (3.9)$$

Note that the *CRPS* is negatively oriented (smaller values are better) and it has the same dimension as the forecasted variable. The *CRPS* is a proper score meaning that it obtains the best expected value when the forecast distribution is equal to the true distribution of probability of the observation. Besides, using proper scoring rules allows the decomposition of the score into the two important attributes of the quality of a forecasting probabilistic model namely resolution and reliability. This permits to understand more precisely the characteristics of the quality of the forecast.

CRPS skill score

In a similar manner, skill scores are used to assess the forecast skill of deterministic forecasts (Coimbra et al., 2013), Pedro et al. (2018) used the *CRPS skill score* (or *CRPSS*) to gauge the quality of their probabilistic forecasting models against a reference method. The *CRPSS* metric (in %) reads as

$$CRPSS = 100 \times \left(1 - \frac{CRPS_m}{CRPS_{ref}} \right), \quad (3.10)$$

where $CRPS_{ref}$ denotes the $CRPS$ of the reference method and $CRPS_m$ refers to the $CRPS$ of the model under evaluation (see Table 3.2.2). A negative value of $CRPSS$ indicates that the probabilistic method fails to outperform the reference model while a positive value of $CRPSS$ means that the forecasting method improves on the reference model. Further, the higher the $CRPSS$, the better the improvement. In this work, and following the recommendations of Doubleday et al. (2020), the raw output of the ECMWF-EPS serves as the reference benchmark model.

Decomposition of the $CRPS$

The decomposition of the $CRPS$ is given by :

$$CRPS = REL_{CRPS} - RES_{CRPS} + UNC_{CRPS}, \quad (3.11)$$

where REL_{CRPS} , RES_{CRPS} and UNC_{CRPS} are respectively the reliability part, the resolution part and the uncertainty part of the $CRPS$. The interested reader is referred to (Lauret et al., 2019) for details regarding the computation of the different components of the $CRPS$.

In addition to reliability and resolution, the uncertainty term accounts for the variability of the observations. It is an indication of the difficulty to forecast the variable of interest and cannot be modified by the forecasting model. It is also worth noting that the uncertainty part UNC corresponds to the score of the climatology. For scores like $CRPS$ that are negatively oriented, the goal of a forecasting model is to minimize (resp. maximize) as much as possible the reliability term (resp. the resolution term). In fact, a forecasting model with a high resolution term means that the model has captured the maximum of the variability present in the data (which variability is measured by the uncertainty term).

3.3.3 Contributions of the statistical moments of the forecast distribution to the $CRPS$

In this study, a new methodology for a better understanding of the skills of a probabilistic forecast in relation with the $CRPS$ is developed. The main idea is to assess separately the contribution of the statistical moments (mean, variance, etc.) of the predictive distributions to the $CRPS$ and consequently to the quality of a probabilistic forecasting model. The principle of the method is to create two virtual forecasts which show the contribution of the statistical moments of the actual forecast to the $CRPS$. Let us illustrate the methodology with 3 forecast PDFs depicted in Figure 3.3.1. \hat{f} represents the actual forecast PDF and \hat{f}_{m_1} and \hat{f}_{m_2} the associated virtual PDF forecasts.

The first virtual forecast \hat{f}_{m_1} is derived from the first moment (mean) of the actual forecast \hat{f} . Let m_1 be the first moment of \hat{f} and δ the Dirac distribution (corresponding to the dotted vertical line in Figure 3.3.1), the PDF of \hat{f}_{m_1} is thereby defined by:

$$\hat{f}_{m_1}(x) \equiv \delta(x - m_1). \quad (3.12)$$

Note that this definition implies that the second, third and further moments of \hat{f}_{m_1} are equal to 0.

The second virtual forecast \hat{f}_{m_2} is given by a Gaussian distribution with first and second moments equal to those of \hat{f} . Let m_2 be the second moment of \hat{f} , \hat{f}_{m_2} is defined as:

$$\hat{f}_{m_2} \sim \mathcal{N}(m_1, m_2). \quad (3.13)$$

Being a Gaussian distribution, the third, fourth and further moments of \hat{f}_{m_2} are equal to 0.

The contribution of the statistical moments of the distribution to the *CRPS* is computed as follows. First, the $CRPS_n$ of each forecast namely $CRPS_{\hat{f},n}$, $CRPS_{\hat{f}_{m_1},n}$ and $CRPS_{\hat{f}_{m_2},n}$ are averaged over the N forecast/observation pairs. This leads to the corresponding values $CRPS$, $CRPS_{m_1}$ and $CRPS_{m_2}$. Second, the difference $G_2 = CRPS_{m_1} - CRPS_{m_2}$ and $G_+ = CRPS_{m_2} - CRPS$ are calculated. Note that one can therefore rewrite the *CRPS* as:

$$CRPS = CRPS_{m_1} - G_2 - G_+. \quad (3.14)$$

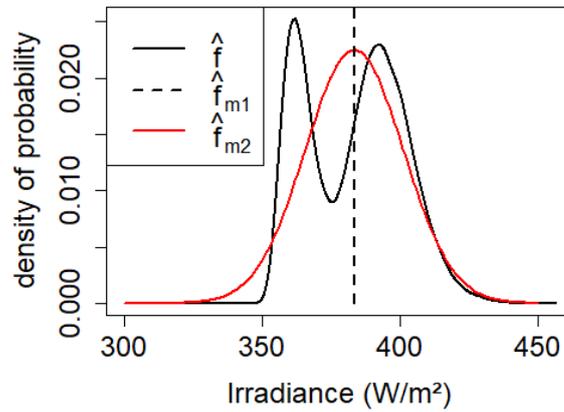


Figure 3.3.1: Illustration of the virtual forecasts \hat{f}_{m_1} and \hat{f}_{m_2} related to the forecast \hat{f} .

Note that the $CRPS_{m_1}$ of the deterministic forecast \hat{f}_{m_1} is actually its Mean Absolute Error (*MAE*)² (see (Hersbach, 2000) for details).

G_2 is the measure of the gain in *CRPS* or equivalently in forecast quality that results from the additional information brought by the second moment of the distribution. G_+ represents the gain resulting from the other statistical moments. G_2 is assumed to be positive. If it is found negative, then the probabilistic forecast has no added value compared to a deterministic forecast. Indeed, the *CRPS* of the probabilistic forecast would be higher than the *CRPS* of the deterministic one ($CRPS_{m_1}$, which is the *MAE*), thus denoting a loss of quality of the probabilistic forecast. On the other hand, G_+ is generally positive. It can be null or negative if the forecast distribution obtains a higher *CRPS* than a Gaussian distribution defined by $\mathcal{N}(m_1, m_2)$. This would indicate that the forecast distribution is less suitable than a Gaussian distribution.

²Similarly to *RMSE*, *MAE* is also a common metric used to assess the accuracy of deterministic forecasts.

In section 3.5.3 below, we propose to present this diagnostic tool under the form of a bar-plot, where $CRPS$, G_+ and G_2 are stacked in this order. G_2 is denoted by the pink part of the bar, G_+ by the green part and $CRPS$ by the blue part. Note that a black line on the top of the blue part is used to better highlight the value of the $CRPS$ and a dotted black line indicates $CRPS_{m1}$ (see Figure 3.5.2). In the following, we refer to this diagnostic tool based on the contribution of the moments of the forecast distributions to the $CRPS$ as “MC-CRPS”.

3.4 Case studies

Six sites are chosen to test the selected models. The first one, Desert Rock, which is part of the SURFRAD network, is located in an arid area. It experiences a high occurrence of clear skies and consequently a very low variability. Two other sites, the airport of Hawaii, where the NREL set up a radiometric network, and Saint-Pierre, which is located on the coastal part of the island La Réunion, are insular sites. Both present a high yearly solar irradiation but also an important variability due to frequent partly cloudy skies. These differences between the two types of sites will permit testing the models under different sky conditions. For an extensive study on the multiple factors that impact the climatology and sky conditions in the specific case of Saint-Pierre and La Réunion, see Badosa et al. (2013) or Kalecinski (2015).

As the aforementioned sites exhibit a similar level of irradiation, three other BSRN sites namely Palaiseau, Tiruvallur and Langley are also considered to test our methodology. The six chosen sites experience different levels of annual solar irradiation and of sky conditions. Thus, this set of sites is representative of the various climates around the world. The main characteristics of these six sites are given in Table 3.4.1. The solar variability, presented in the last line of Table 3.4.1, is defined as the standard deviation of the changes in the clear sky index (Hoff and Perez, 2012).

3.4.1 Measurements

The measured data used in this work are global horizontal irradiance (GHI) time series recorded at the six considered sites. These datasets have been prepared for previous works related to the development and the benchmarking of probabilistic solar forecasts (David et al., 2016; David et al., 2018). They correspond to two years of data divided in a training set (the first year) and test set (the second year). As the ensemble forecasts used here are provided with a 3-hour time step, the recorded time series, initially formatted with a 1-hour granularity, were averaged with a 3-hour time step. A quality check and several test were performed on the recorded GHI time series. The results are given in Appendix 3.A.

	Desert Rock (USA)	Hawaii (USA)	Saint-Pierre (Reunion)
Acronym	DR	HAW	SP
Provider	SURFRAD	NREL	PIMENT
Position	36.6N, 119.0W	21.3N, 158.1W	21.3S, 55.5E
Elevation (m)	1007	11	75
Climate type	Desert	Insular tropic	Insular tropic
Years of record	2012 - 2013	2010-2011	2012 - 2013
Annual solar irradiation (MWh/m^2)	2.105	1.969	2.053
Solar variability 1-h ($\sigma\Delta kt^*_{1hour}$)	0.146	0.209	0.241
	Palaiseau (France)	Tiruvallur (India)	Langley (USA)
Acronym	PAL	TIR	LAN
Provider	BSRN	BSRN	BSRN
Position	48.7N, 2.2E	13.1N, 80.0E	37.1N, 76.4W.
Elevation (m)	156	36	3
Climate type	Mild oceanic	Monsoon	Humid
Years of record	2016-2017	2018-2019	2015-2016
Annual solar irradiation (MWh/m^2)	1.172	1.835	1.685
Solar variability 1-h ($\sigma\Delta kt^*_{1hour}$)	0.281	0.190	0.186

Table 3.4.1: Main characteristics of time series of recorded global horizontal irradiance (GHI) used to test the models.

3.4.2 Forecasts

As mentioned above, the initial day-ahead ensemble forecasts, covering the same period as the measurements, are provided by the European Centre of Medium-Range Weather Forecasts (ECMWF). The EPS is released by ECMWF at 12:00 for the 72 next hours with a 3-hours timestep which allows it to be used for day-ahead scheduling or trading purposes.

3.5 Results

Based on the verification framework proposed by Lauret et al. (2019), the overall performance of the different probabilistic methods is measured by the *CRPS* and the *CRPSS*. Detailed insight in the quality of the models is obtained through the decomposition of the *CRPS* and the new “MC-CRPS” method. Note that this section is dedicated to the presentation of the main results of the study. The next section will be devoted to an in-depth discussion related to the pros and cons of each approach and the added-value brought by the MC-CRPS methodology.

3.5.1 Overall performance of the methods

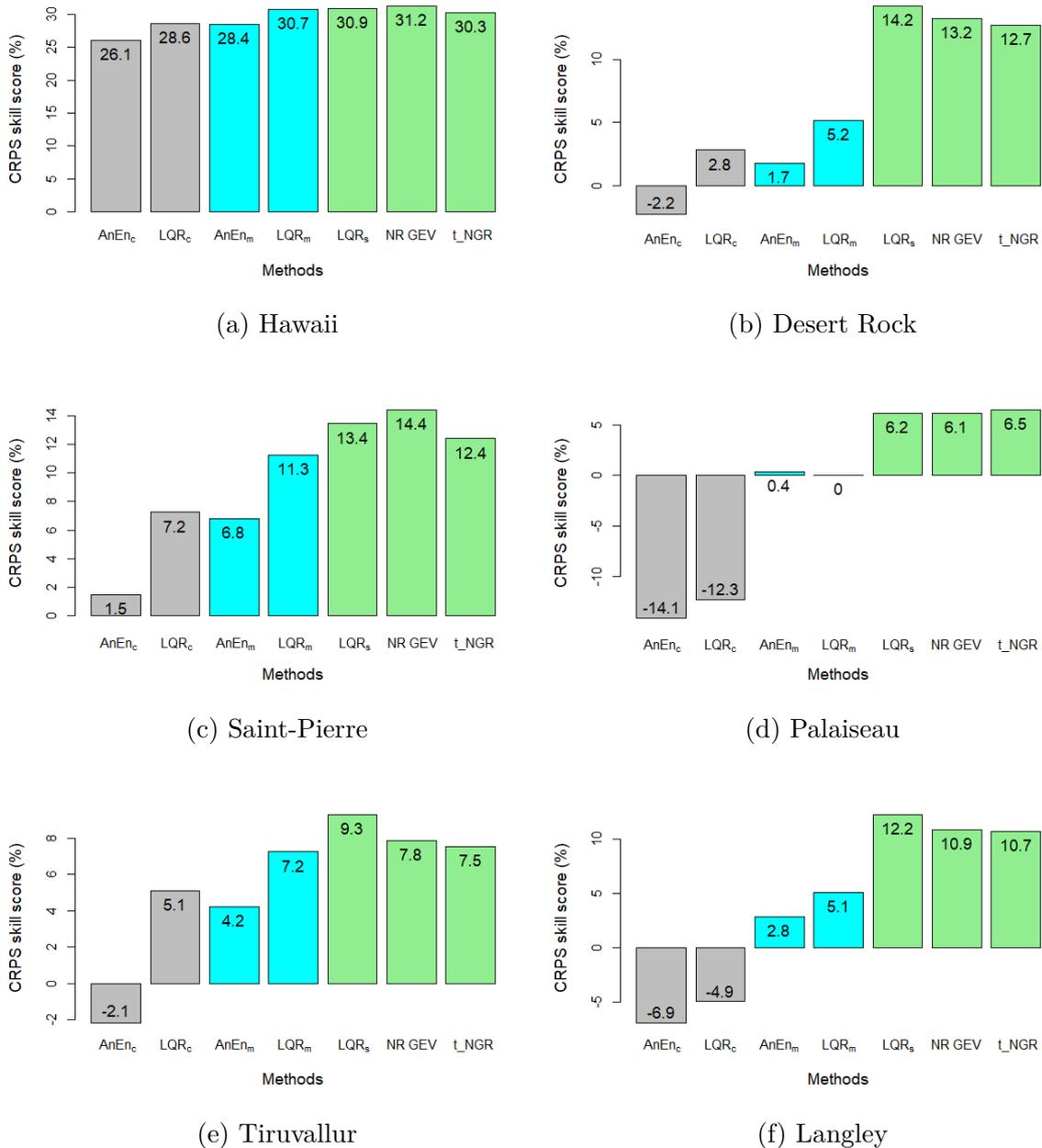


Figure 3.5.1: *CRPS skill score* of each model for the six considered sites. Grey : deterministic-based approach, cyan and green : ensemble-based approach.

Table 3.5.1 lists the *CRPS* obtained by the different methods. However, in order to better highlight the relative merits of each approach, Figure 3.5.1 shows the *CRPS skill scores* of all the forecasting models. Let us recall that positive values of skill scores mean that the model outperforms the reference model (here the raw ECMWF-EPS) while negative values reveal that the quality of the evaluated model is worse than the reference one.

As shown by Figure 3.5.1, regardless the site under study, the highest *CRPS skill scores* are obtained by the ensemble-based approach (represented by the cyan and green

bars). Conversely, except the case of Hawaii, the deterministic-based approach (grey bars) yields lower or even negative skill scores. These negative *CRPSS* values indicate that the deterministic-based models do not always achieve to increase the quality of the raw ensemble forecasts (see for example Palaiseau and Langley).

A deeper look into the performance of the ensemble-based approach shows that models using the mean and the standard deviation of the ensemble members (green bars) exhibit a better forecast skill than models using only the mean of the members (cyan bars) albeit the improvement is less pronounced for Hawaii. Overall, the model with the highest skill score appears to be either LQR_s or NR GEV. Regarding the latter, it may suggest that a judicious choice of the underlying PDF (see Equation 3.6) used by a calibration technique like Nonhomogenous Regression (NR) can further improve the quality of the probabilistic forecasts.

Finally, in order to quantify the relative improvement provided by the ensemble-based approach over the deterministic-based approach, we calculate the gain in *CRPS* based on the *CRPS* values of the best performer of each approach. It appears that the level of improvement is very dependent on the studied site. It is moderate for Hawaii and Tiruvallur (4%), becomes larger for Saint-Pierre (approximately 8%) and quite significant for Desert Rock (approximately 12%), Langley and Palaiseau (approximately 16%).

3.5.2 Detailed insight through the decomposition of the *CRPS*

Table 3.5.1 also provides the decomposition of the *CRPS* into reliability and resolution of the different forecasting methods. As mentioned previously, a forecast should exhibit a small reliability term and a large resolution term. It is worth mentioning first that all models significantly improves the reliability component of the raw EPS forecasts and that the level of improvement strongly depends on the reliability of the initial raw ensemble. Second, it can be noted that the reliability of all calibrated forecasts is fairly comparable. In addition, regardless the site, it appears that, overall, the ensemble-based approach does not significantly improve reliability compared to the deterministic-based approach. Looking in more details, models based on the AnEn technique often appears to generate the most reliable forecasts while the t_NGR model generally provides the less reliable forecasts. Also, in the case of Non homogeneous calibration technique, GEV distributions seem to be more suitable than Gaussian distributions, since NR GEV is slightly more reliable than the t_NGR model.

Regarding the resolution component, it must be noted first that the deterministic-based approach fails to improve the resolution of the raw Ensemble. Conversely, resolution increases with the ensemble-based approach, and particularly when the spread of EPS members is taken as as input of the models i.e. case of the LQR_s, t_NGR and NR GEV models. Finally, one can state that the decomposition of the *CRPS* given in Table 3.5.1 reveals that the difference in quality between the two approaches is mainly explained by the resolution component, whereas reliability is fairly comparable.

	Site	HAW	DR	SP	PAL	TIR	LAN
<i>CRPS</i> (W/m ²)	AnEn _c	50.1	30.1	58.5	44.0	47.8	42.8
	LQR _c	48.4	28.6	55.1	43.3	44.4	42.0
	AnEn _m	48.5	28.9	55.3	38.4	44.9	38.9
	LQR _m	46.9	27.9	52.7	38.6	43.4	38.0
	raw Ensemble	67.7	29.4	59.4	38.6	46.8	40.0
	LQR _s	46.8	25.2	51.4	36.2	42.5	35.2
	t_NGR	47.2	25.7	52.0	36.2	43.3	35.8
	NR GEV	46.6	25.5	50.8	36.2	43.2	35.7
Reliability (W/m ²)	AnEn _c	4.2	4.8	6.6	4.9	7.2	4.8
	LQR _c	4.4	5.3	7.1	5.4	6.7	5.3
	AnEn _m	4.1	4.7	6.2	4.9	7.9	4.5
	LQR _m	4.4	5.7	7.0	5.7	8.2	5.0
	raw Ensemble	23.2	8.4	13.4	7.5	11.5	8.2
	LQR _s	4.5	5.9	7.6	5.3	8.2	5.4
	t_NGR	4.7	6.5	8.4	5.4	8.0	5.7
	NR GEV	4.1	6.2	7.2	5.4	7.8	5.8
Resolution (W/m ²)	AnEn _c	111.9	149.7	120.8	87.4	120.4	116.4
	LQR _c	113.9	151.7	124.7	88.6	123.3	117.7
	AnEn _m	113.4	150.8	123.5	93.0	124.0	120.0
	LQR _m	115.3	152.8	127.0	93.6	125.8	121.4
	raw Ensemble	113.3	154.0	126.7	95.4	125.7	122.5
	LQR _s	115.5	155.6	128.9	95.6	126.8	124.7
	t_NGR	115.3	155.7	129.0	95.8	125.8	124.3
	NR GEV	115.3	155.6	129.1	95.6	125.7	124.5
Uncertainty (W/m ²)	All Models	157.8	175.0	172.7	126.5	161.0	154.4

Table 3.5.1: *CRPS* and its components reliability, resolution and uncertainty of all considered models for the 6 sites. Grey : deterministic-based approach, cyan : model using the mean of the members as predictor (ensemble-based approach), green : model using the mean and spread of the members as predictors (ensemble-based approach). Red values indicate the worst *CRPS*s while the black bold ones show the best *CRPS*s.

3.5.3 Detailed insight through the CRPS Moments-Contributions

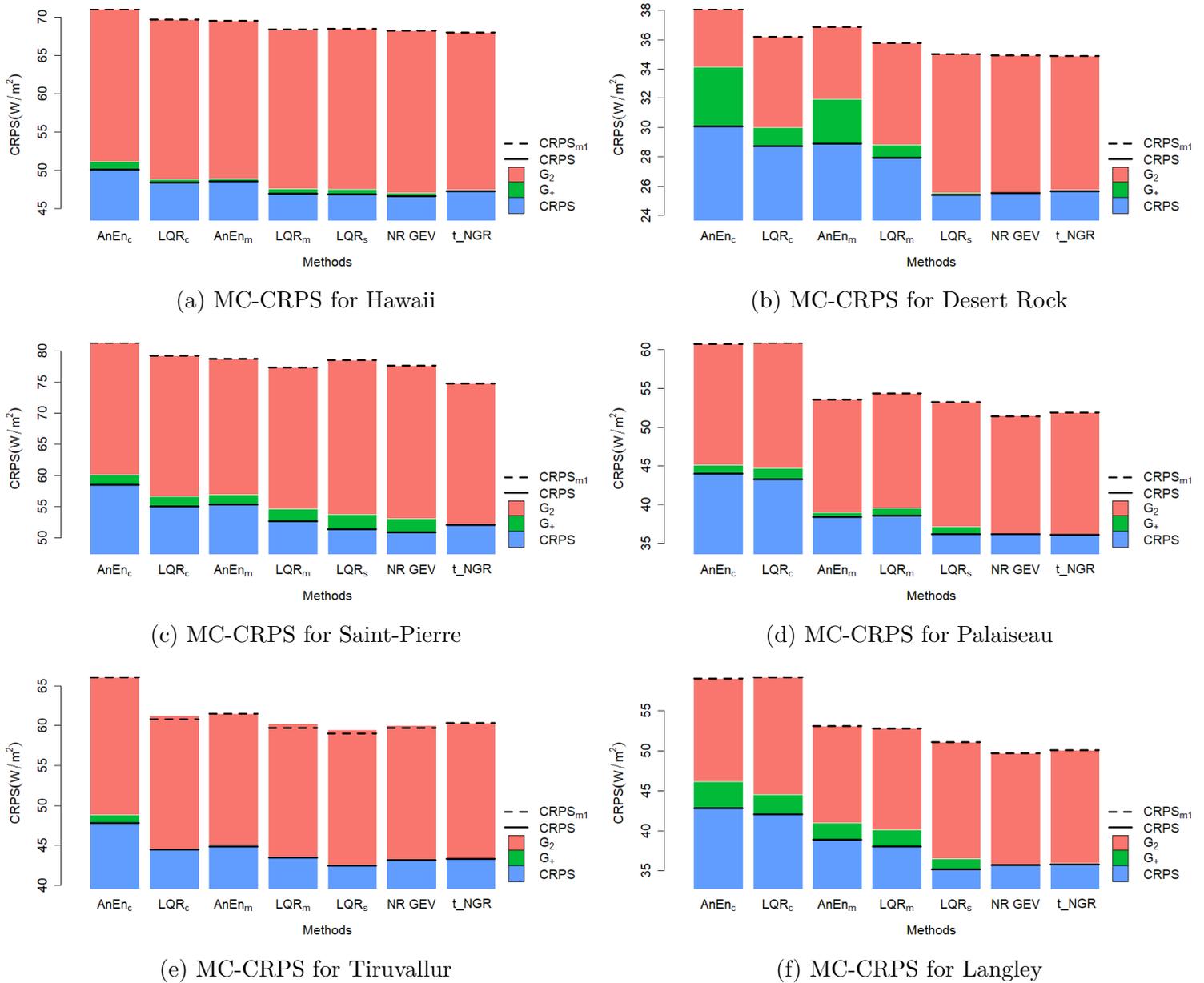


Figure 3.5.2: MC-CRPS of the six sites and all forecasting models. A black line is used to better highlight the value of the $CRPS$ and a dotted black line indicates the value of $CRPS_{m1}$.

Figure 3.5.2 shows the results of the MC-CRPS introduced in section 3.3.3. As seen, the final $CRPS$ values, of the forecasting models occur to be dependent on their respective $CRPS_{m1}$ values. In particular, models from the ensemble-based approach appear to have better $CRPS_{m1}$ than models from the deterministic-based approach (see for instance the case of Langley). This means that the aggregation of members improves the estimation of the first moment. Among the ensemble-based models, except for the cases of Hawaii and Tiruvallur, the superiority of the LQR_s , NR GEV and t_NGR models using the mean

and standard deviation of the ensemble members can be mainly explained by a greater contribution of G_2 . Thus the spread of EPS members is effective and improves more importantly G_2 than $CRPS_{m1}$. Further, the best performer among the three aforementioned models is finally determined by G_+ . This highlights the importance of the choice of the distribution in the non homogenous regression calibration framework.

For example, overall, NR GEV performs better than the t_NGR model because of G_+ . Let us stress that the choice of strictly truncated Gaussian distributions in the implementation of a NGR technique forces G_+ to be very close to 0 in the MC-CRPS. Hence, the benefits of GEV distributions compared to Gaussian distributions are highlighted by the MC-CRPS method.

3.6 Discussion

In this section, we try to give more clues regarding the merits of each proposed approach. Also, a discussion related to the advantages brought by the MC-CRPS is proposed.

3.6.1 Deterministic-based approach versus ensemble-based approach

Let us recall that the deterministic-based approach uses a unique deterministic predictor while the ensemble-based approach makes use of the information conveyed by the ensemble. Therefore, the main weakness of the deterministic-based approach is the lack of information feeding the models. Since the distribution needs to be completely determined from one single deterministic predictor, the spread and the possible skewness and kurtosis of the forecasting distribution need to be only inferred from this single predictor. Conversely, the benefits gained from the multiplicity of predictors provided by the ensemble-based approach need to be significant to justify the computation of the EPS. Two types of benefits can be discussed.

First, the aggregation of predictors leads to a better estimation of the first moment. This is visible in Figure 3.5.2 where models issued from the ensemble-based approach get better $CRPS_{m1}$ than models from the deterministic-based approach. It is clear that a gain in the estimation of the first moment can be obtained by the substitution of the control member by the mean of all members.

Second, regarding the determination of the second moment, the uncertainty is already carried by the level of forecasting of the mean of the EPS members. These variables are dependent, as shown in Appendix 3.B (the standard deviations of the observations clearly depends on the level of forecasting). Hence, using the spread of the members of EPS as input of the forecasting models can only be justified if it brings an extra-information on the uncertainty. It is assumed that the spread of the members is higher if the uncertainty is so. Indeed it indicates if slight errors in the initial conditions could lead to great differences in the final state of the atmosphere.

Thus, it appears necessary to investigate on the quantity of information actually provided by the spread of the members. In order to do this, the correlation between the standard deviation of the observations and the spread of the members has been studied. This has been made for a fixed level of forecasting, in order to remove the dependency between uncertainty and level of forecasting. Then an average over all levels of forecasting

has been calculated to produce Figure 3.6.1. This kind of plot is of great utility to know the added value of the standard deviation of the EPS forecast members. If the dependence between the spread of the members and the uncertainty of the forecast for a fixed level of forecasting is strong, then a large improvement can be expected for calibration models using the spread of the members as an input, compared to simpler models.

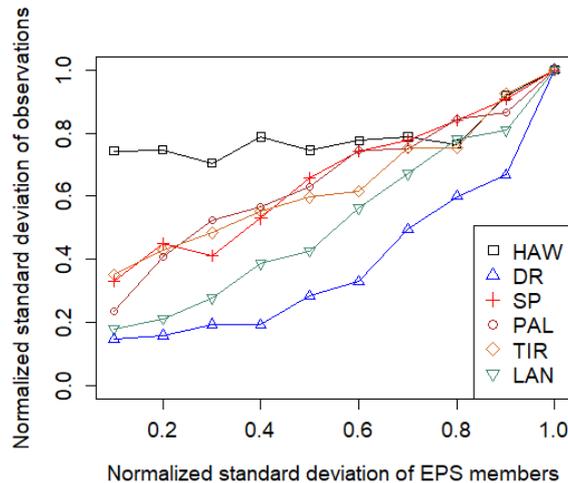


Figure 3.6.1: Standard deviation of observations vs. standard deviation of the EPS members (raw ECMWF ensembles). Normalization of the standard deviation has been done by dividing the standard deviations by the maximum of the standard deviation for each site.

As shown by Figure 3.6.1, the amount of new information given by the spread of the members is very dependent on the studied site. When for Hawaii, the correlation between the standard deviation of the observations and the spread of the members is almost null, it is quite significant for the other sites and especially for Langley and Desert Rock. A link can be established between this finding and Figure 3.5.1 which shows that the success of taking into account the spread of members in the forecasting models depends on the site. It is clearly less valuable in Hawaii than in other sites, and it is particularly successful in Desert Rock and Langley. It is also consistent with Figure 3.5.2 where G_2 is significantly higher in Desert Rock for LQR_s, t_NGR and NR GEV models.

3.6.2 Discussion related to CRPS Moments-Contributions

In order to consolidate the results obtained in Figure 3.5.2, a complete analysis of the statistical moments of the probability distributions produced by the forecasting methods has been conducted. This kind of study is traditionally done to assess the strengths and weaknesses of a forecasting model. Although the deterministic measure of a statistical moment is not a proper scoring rule, it is of great interest to use it to understand the behaviour of the forecasting models.

First, an evaluation of the accuracy of the first moment has been conducted. A good forecasting model should have the ability to give a mean value of the forecasting distributions as close as possible to the mean of the observation values. A measure of this ability can be obtained by calculating the Root Mean Square Error (*RMSE*) or Mean

Absolute Error (MAE) of the mean of the forecasting distributions. In this study, the MAE has been chosen as it is exactly the definition of $CRPS_{m,1}$ introduced in section 3.3.3 (see Hersbach (2000) for details). Figure 3.5.2 gives therefore the results related to the accuracy of the first moment of the distributions.

Second, a probabilistic forecast also provides an estimation on the level of uncertainty, which is reflected by the spread of the forecasting distribution (i.e. the second statistical moment). Some works have been specifically dedicated to the assessment of the accuracy of the spread of the predictive distributions. Among others, one can cite the studies related to the spread-skill relationship (see Whitaker and Lough (1998) or Hopson (2014)). These works are guided by the idea that the variance of a probabilistic forecast should be larger if the uncertainty of the forecast is so. Fortin et al. (2014) proposed a criterion for the evaluation of the accuracy of the second moment of the distributions. This criterion is based on the fact that statistical consistency requires that the spread of the forecasting distributions should be equal to the $RMSE$ of the mean of the forecast. Following Fortin et al. (2014), spread is calculated as the square root of the mean of the variances of the forecasting distributions. The accuracy of the second moment is therefore measured by calculating the $RMSE$ of the differences between spread and $RMSE$ of the mean of the distributions (i.e. $RMSE_M$). Figure 3.6.2 plots the $RMSE$ of the difference ($spread - RMSE_M$), computed over the evaluation period.

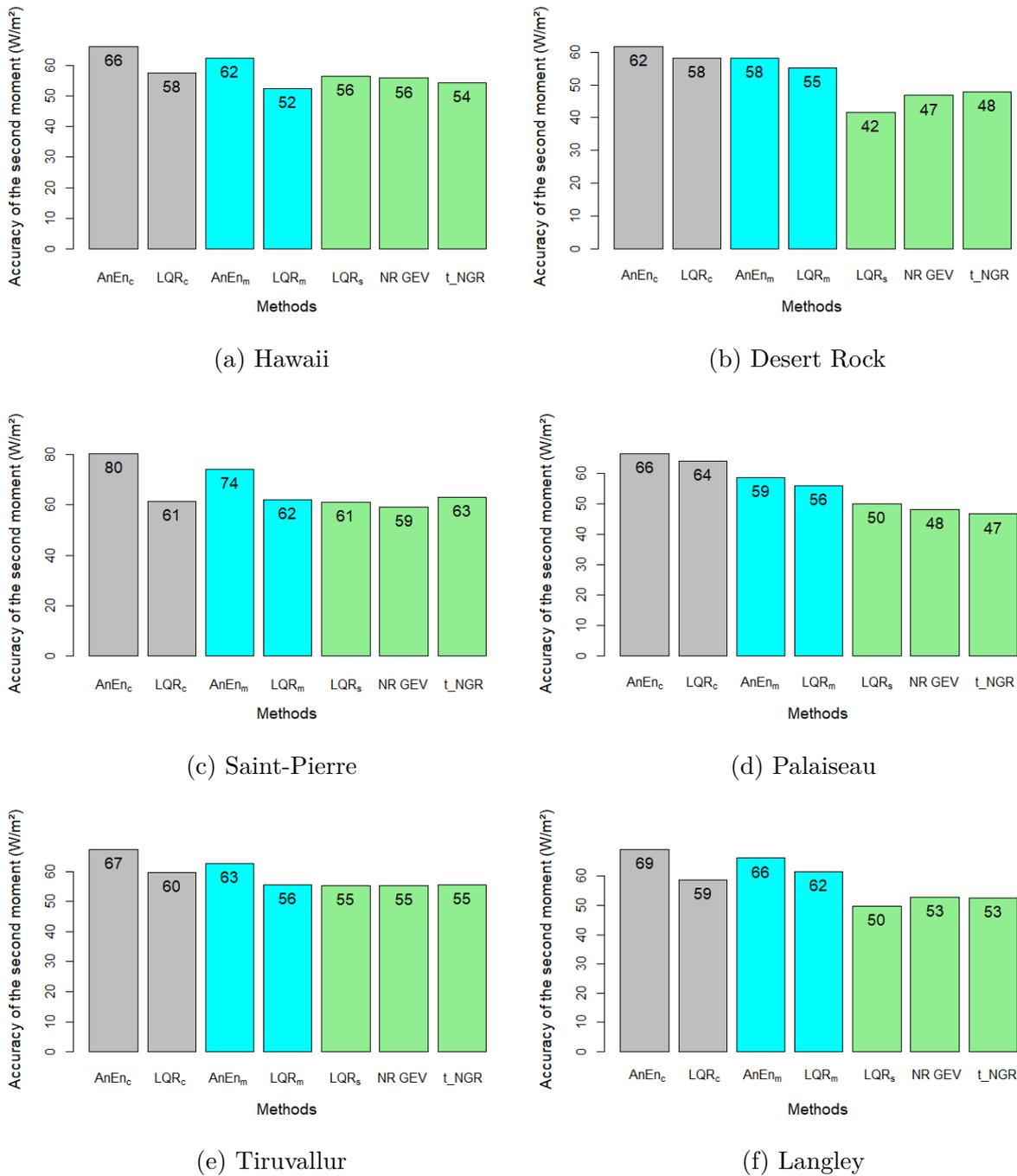
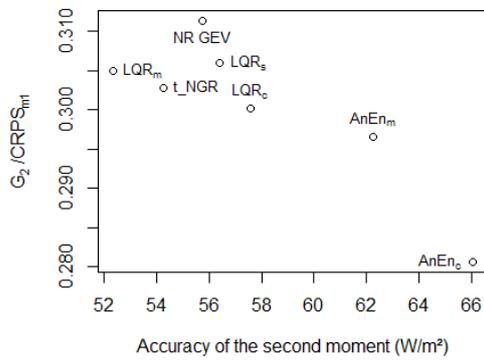


Figure 3.6.2: Accuracy of the second moment for the six studied sites and all forecasting models.

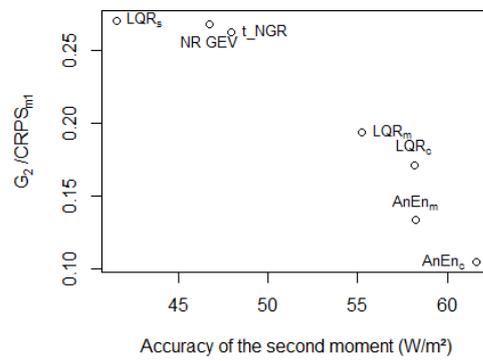
Conversely to the first moment, the accuracy of the second moment gradually improves when the information taken by the forecasting model is more complete. Using the mean of members instead of the control member increases the second moment accuracy. Taking into account the spread of the EPS improves further the accuracy by approximately the same extent (except for Hawaii, for the reasons discussed in section 3.6.1). Nevertheless, this improvement depends on the site. As shown by Figure 3.6.2, the accuracy of the second moment for Hawaii is almost equal for each model. It is consistent with the results depicted in Figure 3.6.1, showing that the information of the second moment of

the EPS distribution in Langley and Desert Rock is the most valuable , as opposed to the information of Hawaii EPS distribution.

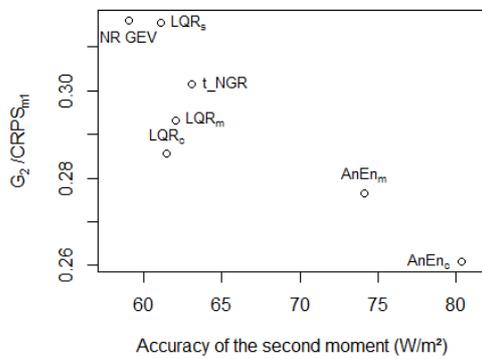
The accuracy of the second moment can be linked to the gain G_2 introduced in the MC-CRPS section (see section 3.3.3). The correlation between these two values is highlighted in Figure 3.6.3, which shows the ratio $\frac{G_2}{CRPS_{m1}}$ versus the accuracy of the second moment.



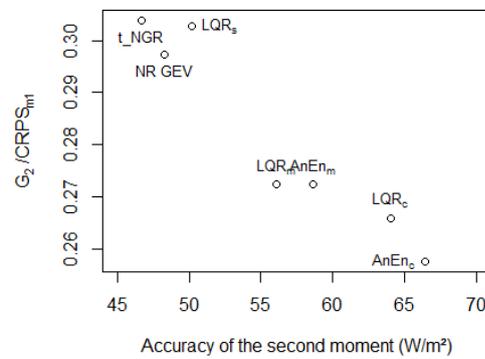
(a) Hawaii



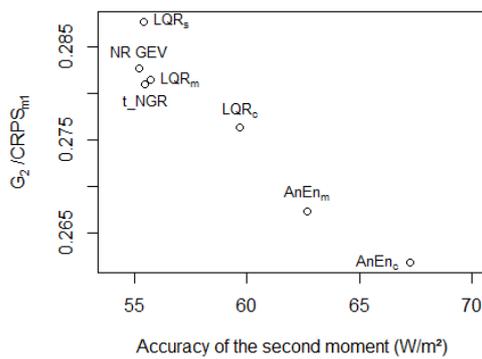
(b) Desert Rock



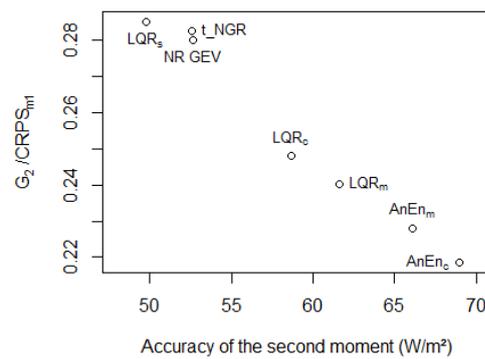
(c) Saint-Pierre



(d) Palaiseau



(e) Tiruvallur



(f) Langley

Figure 3.6.3: Link between G_2 and the accuracy of the second moment.

To sum up, the great advantage of the MC-CRPS is to reconcile the score of a probabilistic forecasting model and the explanation of its performance by examining the accuracy of the moment-based distributions.

Moreover, the link between the calibration of the moments and the score is highlighted,

because the contribution of the accuracy of the moments to the score is quantified. Here, in the proposed new diagnostic tool MC-CRPS, the accuracy of the statistical moments of the forecasting distributions is quantified by the proper score itself. This diagnostic tool is complementary of the decomposition discussed in section 3.3.2, i.e. the reliability and resolution of \hat{f}_{m1} and \hat{f}_{m2} can also be computed and studied. The MC-CRPS diagnostic tool also highlights the benefits of probabilistic forecasting, as the comparison between $CRPS_{m1}$ and $CRPS$ provides a measure of the quality difference between deterministic and probabilistic forecasting.

3.7 Conclusions

Based on the two types of forecasts i.e deterministic or ensemble forecast (denoted by the term EPS for ensemble prediction system) issued by the meteorological centre ECWMF, two approaches for generating day-ahead solar irradiance probabilistic forecasts were proposed. The first approach creates probabilistic forecasts from the deterministic day-ahead GHI predictor while the second one generates probabilistic forecasts from the calibration of the EPS or from information inferred from the EPS.

The goal of this work was to quantify the possible added-value of the EPS on the quality of the forecasts. Six sites experiencing different sky conditions were chosen for the appraisal of the different probabilistic models. Quality of the different probabilistic models have been evaluated with common diagnostic tools such as the $CRPS$ and its decomposition. A new diagnostic tool called MC-CRPS has also been introduced. It consists in the measure of the contribution of each statistical moment of the forecasting distributions to the $CRPS$.

Overall, models adopting the ensemble-based approach have been found to issue probabilistic forecasts with better quality than the ones based on the deterministic-based approach. The gain in quality, based on the $CRPS$ metric, ranges from 4 % up to 16 %.

One other important contribution of this work is the new diagnostic tool related to the $CRPS$ based on the moments of the ensemble distribution called MC-CRPS. This MC-CRPS tool allowed to identify two characteristics of EPS that have an impact on the quality of probabilistic forecasts. First, the aggregation of deterministic predictors of the ensemble leads to an improvement of the estimation of the first moment and thus, raises the overall quality of a probabilistic forecast. Second, the spread of the EPS members turns to be a good predictor that permits to enhance the estimation of the second moment of the forecasting distributions. Finally, in terms of forecast quality, it can be concluded that using an EPS (which requires high computing capacities) to produce day-ahead GHI probabilistic forecasts should be favored compared to a deterministic (less demanding) approach. This work opens the way to the assessment of the forecast value of each approach i.e. the benefit (economical or others) gained from the use of these probabilistic forecasts in an operational context.

Appendices

3.A Data quality check

A quality check has been conducted for the observation data of each of the six studied sites. As the decomposition of irradiance into diffuse and direct has not been measured, the exhaustive set of BSRN recommended quality checks could not be conducted (see Long and Dutton (2010)), but only the first plot. It consists in the plot of measured irradiance versus solar zenith angle. The rarely reached limit is plotted in dashed line and the physical possible limit is plotted in solid line. The second check is a frequency histogram of the clear-sky index (k_t^*) for each site. k_t^* is defined as:

$$k_t^* = \frac{\text{GHI}}{\text{GHI}_{\text{CS}}}, \quad (3.15)$$

where GHI_{CS} , the clear-sky GHI, is calculated with the Bird clear-sky model (Bird and Hulstrom, 1981). The maximum of the observed frequency is supposed to be at $k_t^* = 1$. The third check is a plot of the k_t^* , only for clear-sky days. The morning data is reported by black dots and afternoon data by red dots. From this plot, it is possible to see if clear-sky irradiances are well-reported by the measurement data. If not, the line drawn by the dots is not straight. To extract clear-sky days from the data, the process proposed in Badosa et al. (2013) has been followed. The last figure is a plot of the k_t^* for each hour and day of the year. It allows to detect if systematical biases exist at some days/hours of the year. It also allows to easily detect missing data.

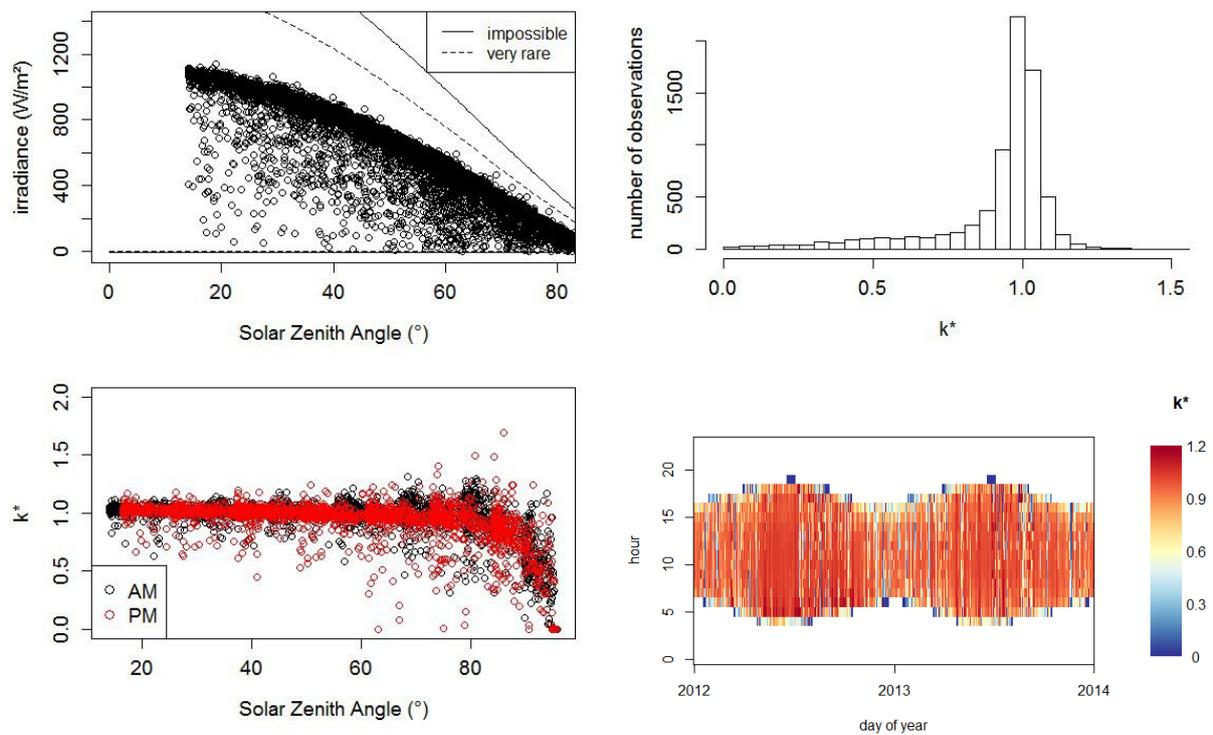


Figure 3.A.1: Quality check for Desert Rock.

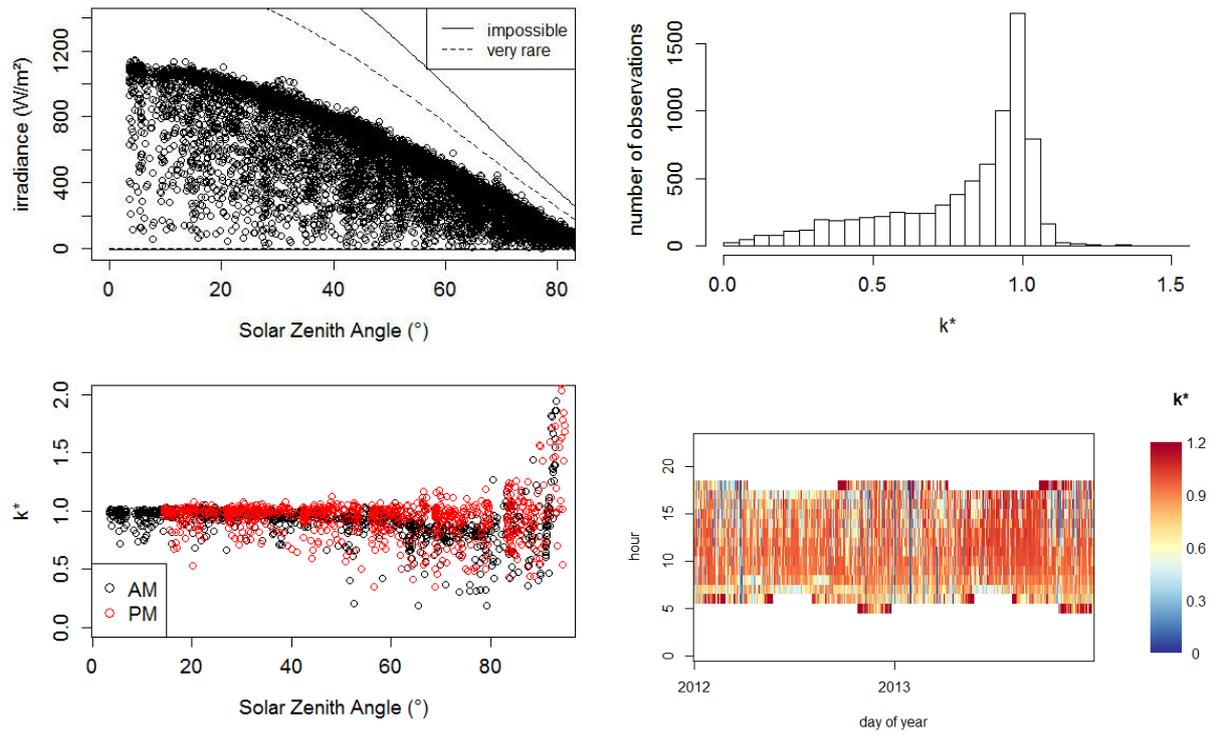


Figure 3.A.2: Quality check for Saint-Pierre.

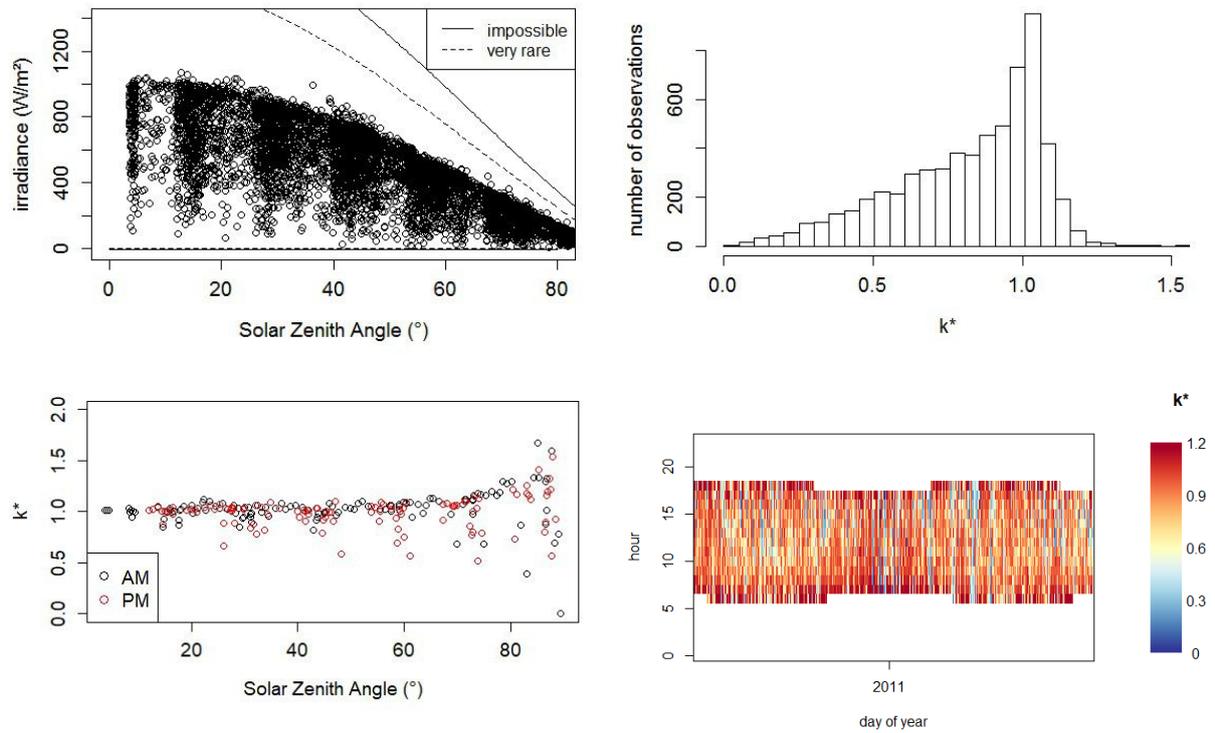


Figure 3.A.3: Quality check for Hawaii.

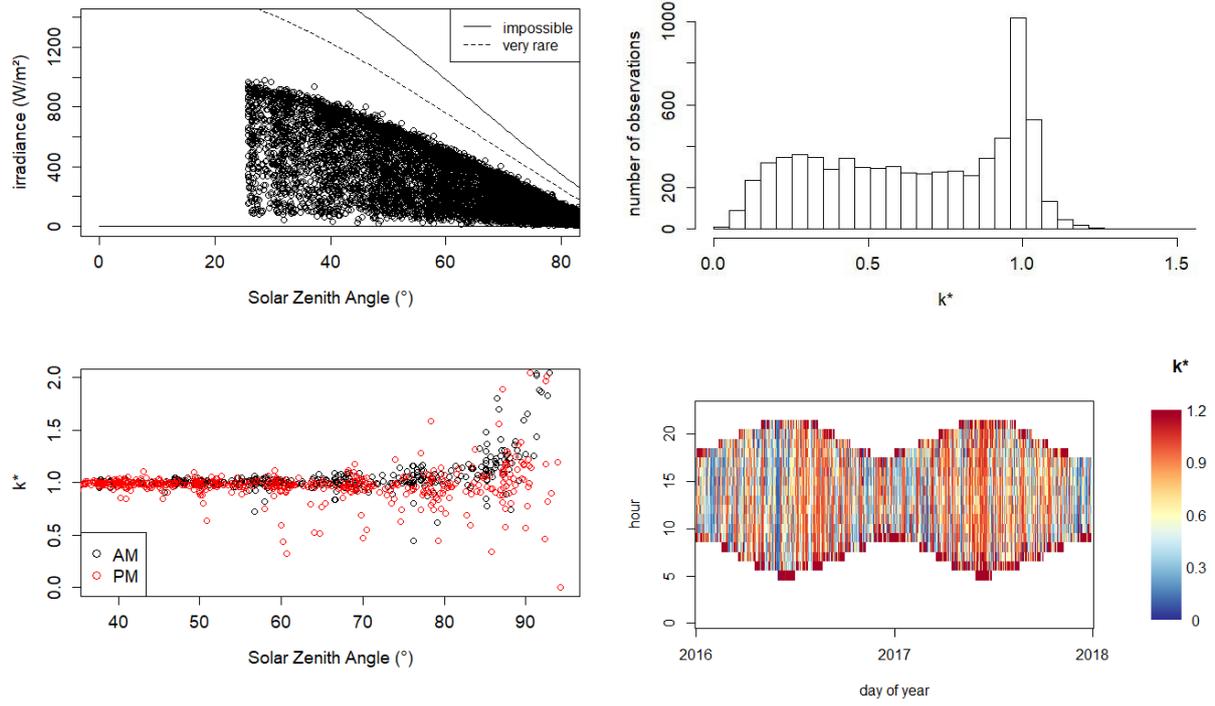


Figure 3.A.4: Quality check for Palaiseau.

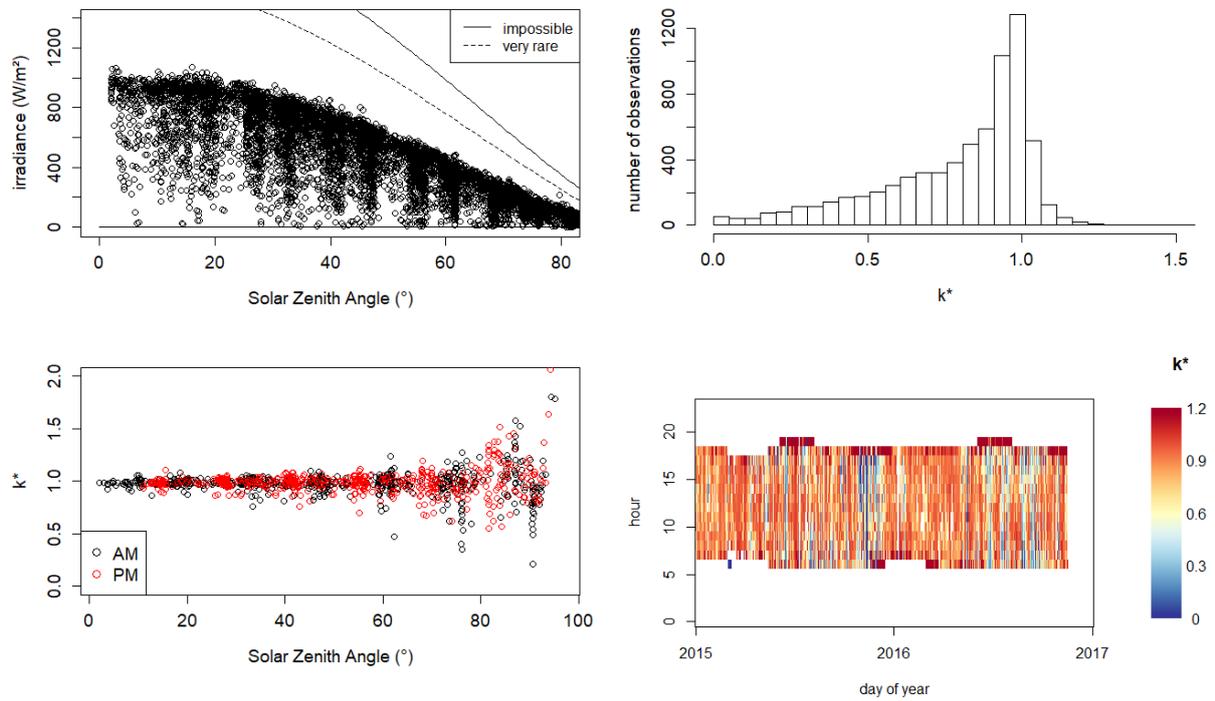


Figure 3.A.5: Quality check for Tiruvallur.

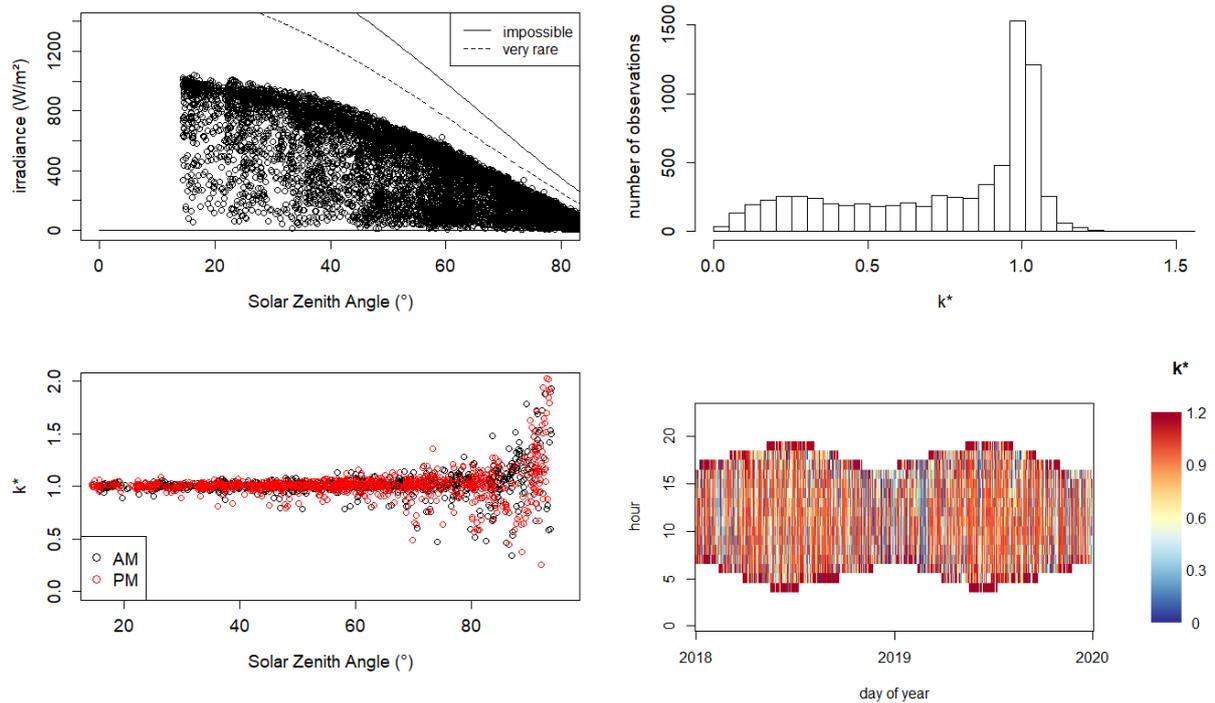


Figure 3.A.6: Quality check for Langley.

No major issues have been detected concerning the six studied sites. For some sites (Tiruvallur, Langley, Saint-Pierre), it is possible to guess that some reflexions occur for extreme hours and some seasons. This leads to the phenomenon of overirradiance where k_t^* can easily reach a value of 4.

3.B Bias and standard deviation of EPS members distribution and observations for the six sites

The definition of the probabilistic forecasts defined by the uniform construction from raw ensembles as presented in section 3.2 is often under-dispersive, and consequently obtains poor scores. The associated rank histograms usually get characteristic U-shapes, with overpopulated extreme ranks. In this section, we attempt to demonstrate why a calibration procedure is needed for these raw forecasts. To this end, a comparison between members distributions and observation distributions depending on the level of forecasting has been conducted for the 2 first statistical moments. These plots show clearly under-dispersive raw ensembles. The standard deviations need to be corrected. The discrepancy between distributions of members and observations indicates a statistical inconsistency between observations and forecasts, and therefore a bad reliability, and justifies the use of calibration models.

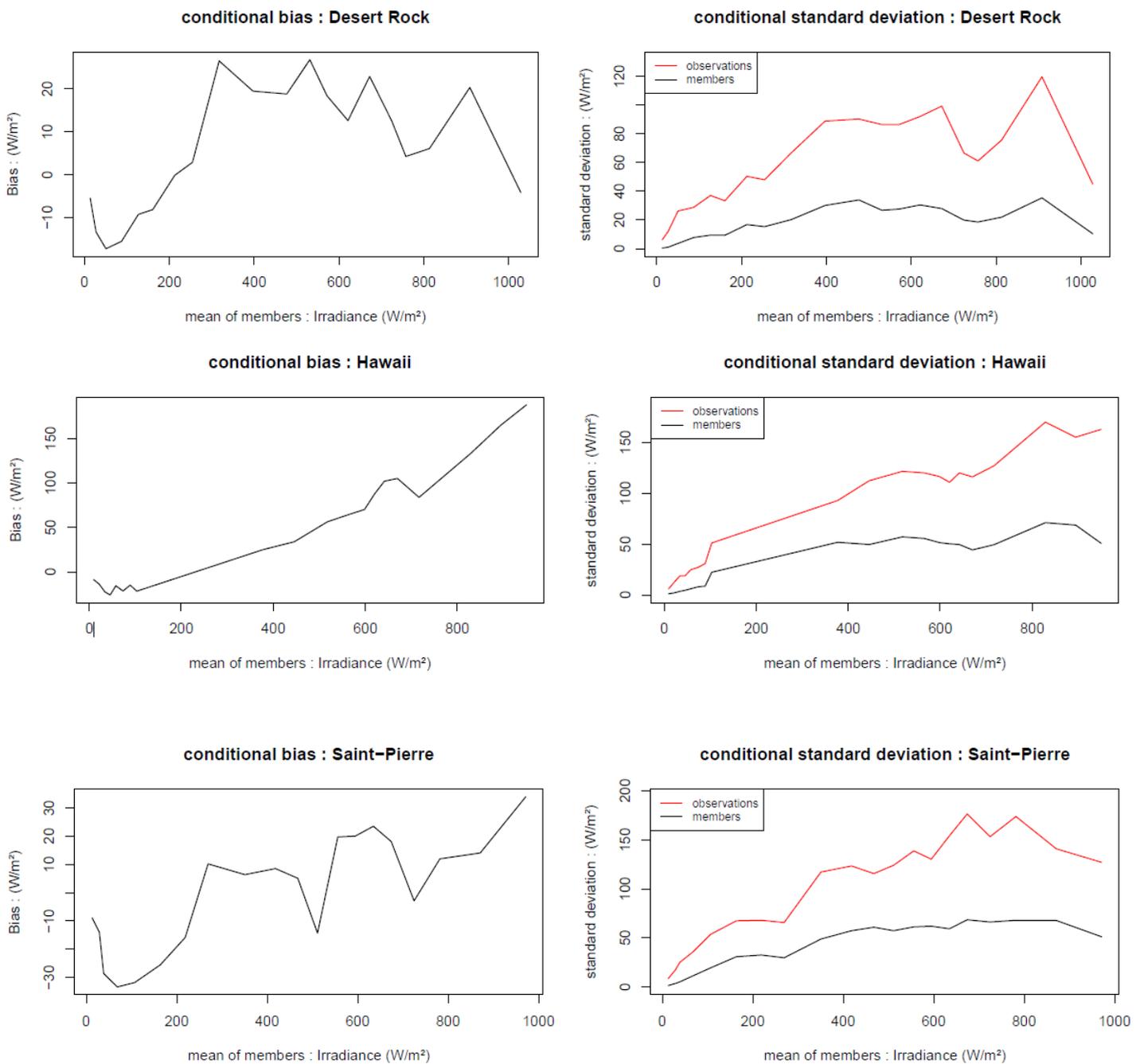


Figure 3.B.1: Bias and standard deviation of EPS members distribution and observations for Desert Rock, Hawaii and Saint-Pierre, depending on the level of forecasting.

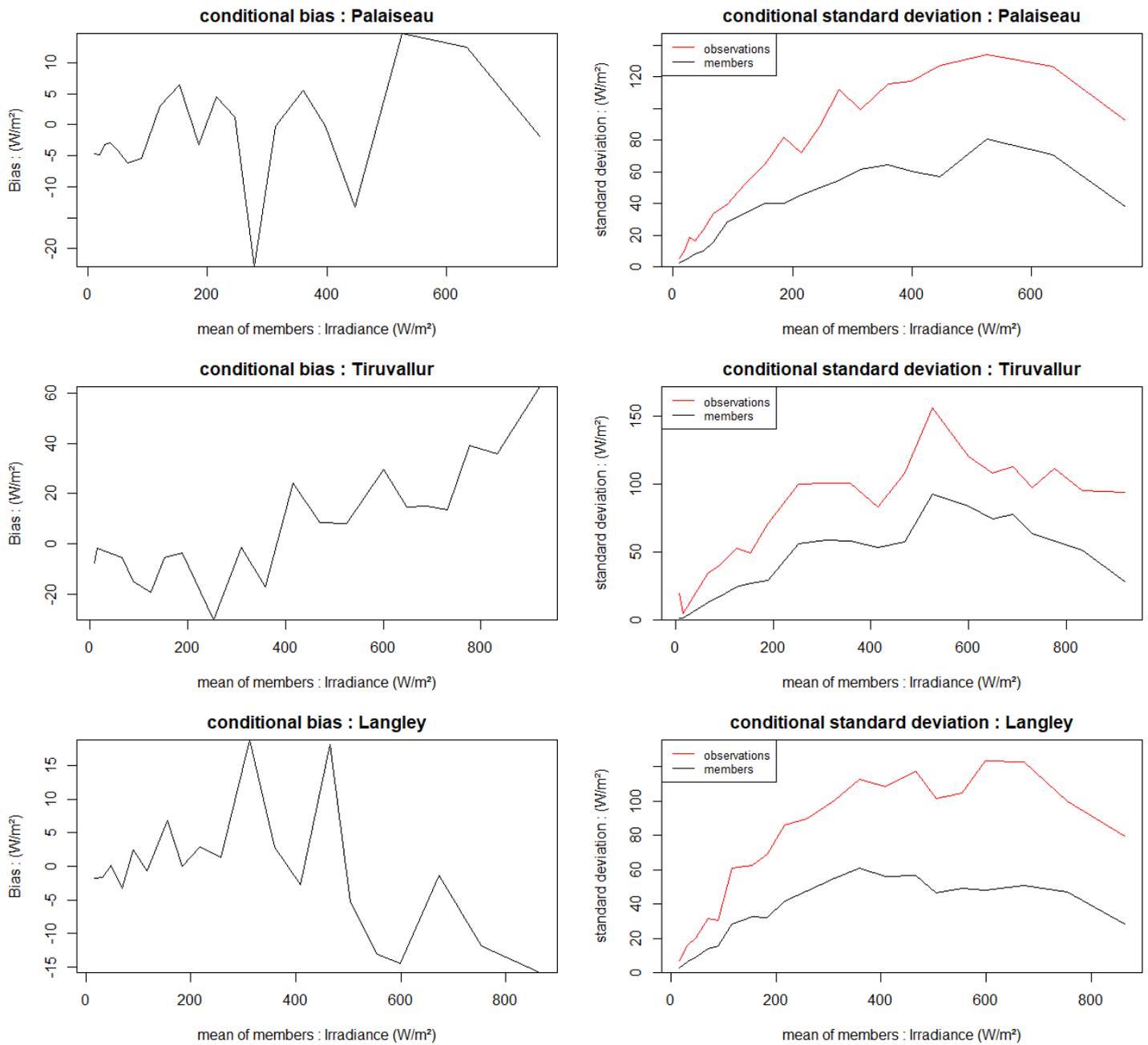
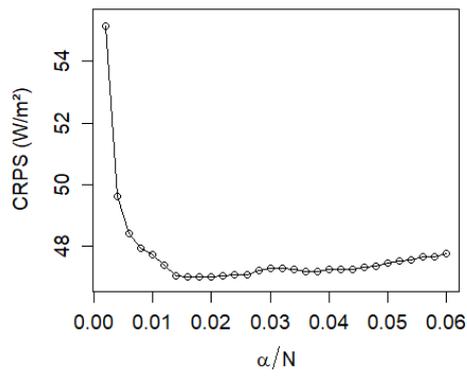


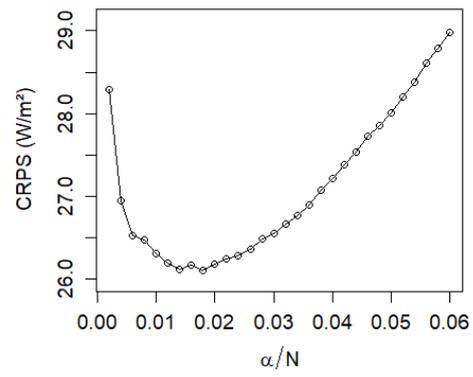
Figure 3.B.2: Bias and standard deviation of EPS members distribution and observations for Palaiseau, Tiruvallur and Langley, depending on the level of forecasting.

3.C Selection of the optimal α

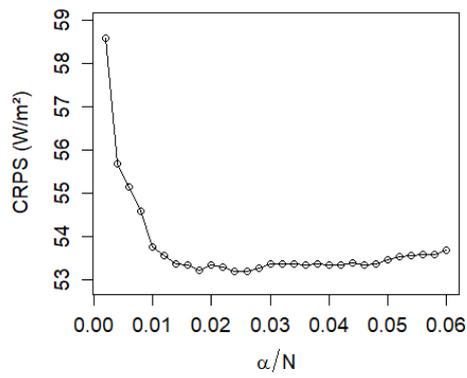
Figure 3.C.1 presents the results related to the optimal selection of the parameter α . As shown by Figure 3.C.1, regardless of the site under study, the optimal value corresponds to the minimum of the *CRPS* calculated on the training evaluation set.



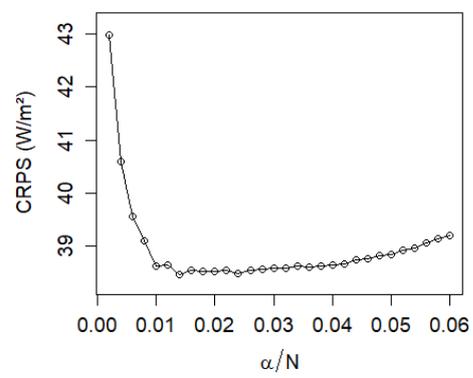
(a) Hawaii



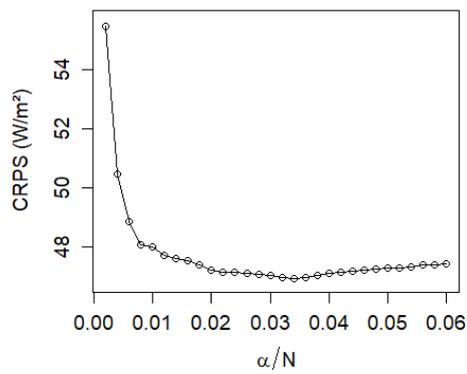
(b) Desert Rock



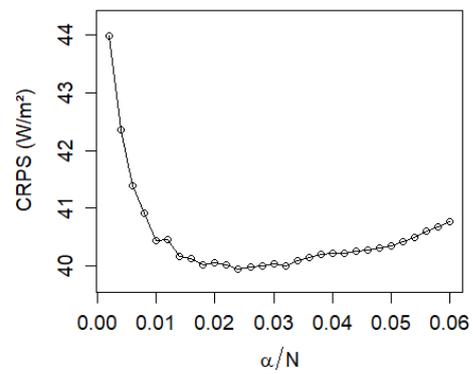
(c) Saint-Pierre



(d) Palaiseau



(e) Tiruvallur



(f) Langley

Figure 3.C.1: Determination of α . The optimal value corresponds to the minimum of the CRPS.

3.8 Compléments

3.8.1 CRPS skill scores par rapport au modèle CSD-CLIM

Dans cette partie, nous tirons profit du modèle de référence climatologique CSD-CLIM pour calculer les *CRPS skill scores* des différents modèles de prévision présentés dans ce chapitre. Les résultats sont présentés sur la figure 3.8.1.

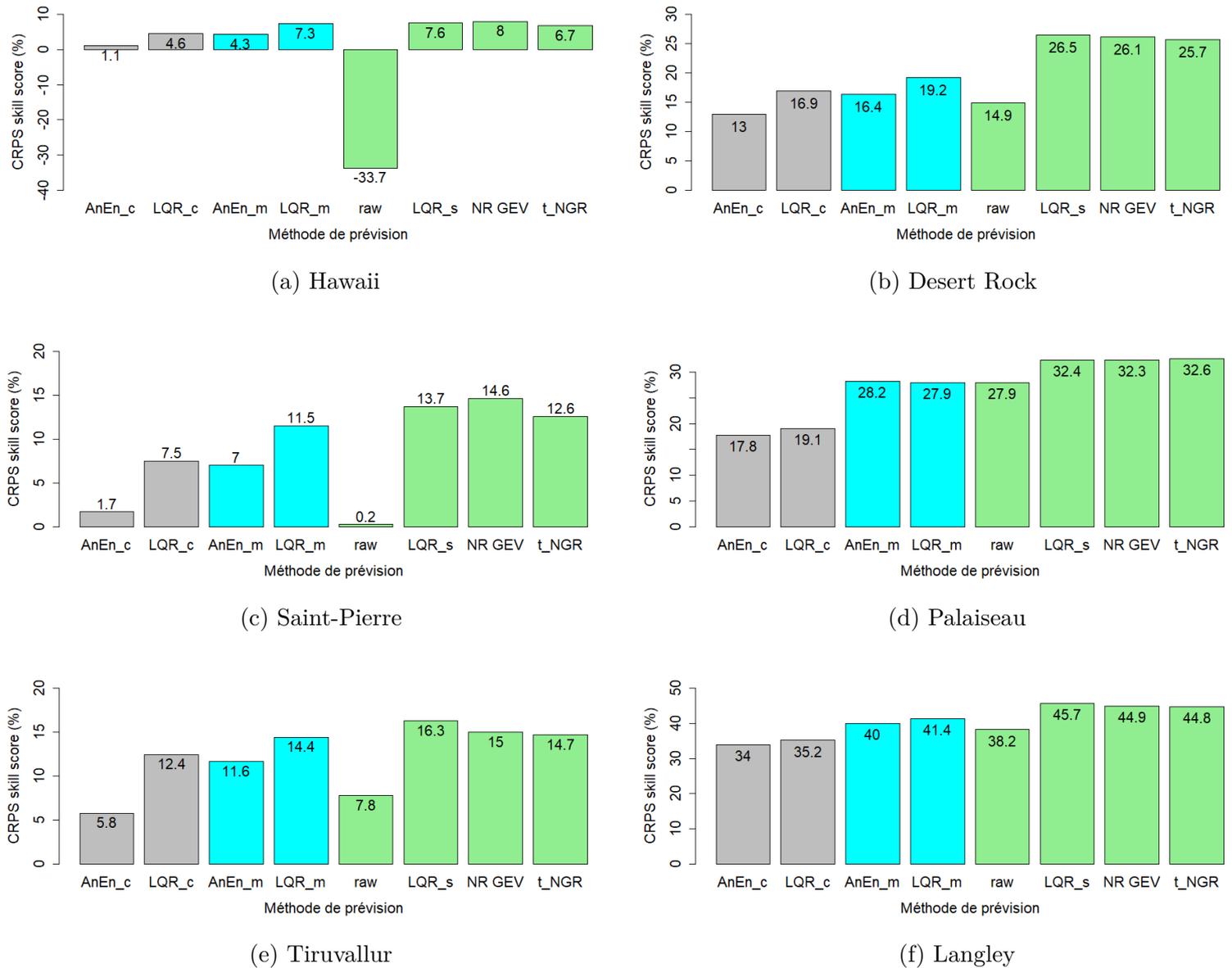


Figure 3.8.1: *CRPS skill scores* des 7 modèles de prévision et de l'ensemble brut calculés par rapport au modèle de référence CSD-CLIM.

Ces résultats peuvent être rapprochés de la figure 3.5.1. Cette dernière présente des *CRPSS* calculés par rapport à un modèle de référence appelé “raw ensemble” dans la terminologie proposée par Doubleday et al. (2020), c’est-à-dire obtenu à partir des prévisions d’ensemble brutes fournies par ECMWF. Ici, les *CRPS skill scores* sont calculés par rapport au modèle de référence climatologique CSD-CLIM introduit dans ce chapitre.

Il est notable que les performances relatives de ces deux modèles de référence sont extrêmement variables. À Hawaii, l’ensemble brut a un *CRPSS* de -33.7% alors qu’il est de 38.2% à Langley. Les *CRPSS* des différents modèles de prévision sont assez liés. Si la corrélation entre les *CRPSS* des modèles de calibration et celui de l’ensemble brut semble faible, les *CRPSS* des sites pour lesquels l’ensemble brut ne parvient pas à avoir une meilleure performance que CSD-CLIM (Hawaii, Saint-Pierre) sont tout de même moins élevés en moyenne que ceux des autres sites. En particulier, les 2 sites pour lesquels la prévision d’ensemble brute est moins performante que CSD-CLIM sont les 2 sites pour lesquels l’écart-type des observations en fonction du niveau de prévision atteint les valeurs les plus élevées (cf. Annexe 3.B).

Pour tous les sites, tous les modèles de calibration sont meilleurs que la méthode de référence climatologique, ce qui souligne l’intérêt de l’utilisation de la prévision d’ensemble pour la production de prévisions probabilistes, et la nécessité de sa calibration. En effet, le modèle CSD-CLIM est le modèle de référence climatologique tirant le meilleur parti de l’information climatologique d’un site (cf. chapitre 2). Cela montre que l’information supplémentaire (par rapport à l’information qu’on peut tirer de la climatologie) apportée par la prévision d’ensemble apporte un gain de qualité, pour autant qu’elle soit correctement traitée par un modèle de calibration. Les modèles “Ensemble-based” sont les meilleurs, tout en présentant des scores assez équivalents.

Cependant, il semble que l’intérêt de l’utilisation des prévisions d’ensemble soit très variable, car les *CRPS skill scores* sont eux-mêmes très variables d’un site à l’autre.

3.8.2 Gains effectifs de qualité obtenus grâce à l’approche probabiliste

Dans ce chapitre, les gains obtenus grâce à l’approche ensembliste ont été mesurés dans la partie 3.5.1. Dans cette partie, on souhaite aussi mesurer les gains dus spécifiquement à l’utilisation de prévisions probabilistes. La figure 3.8.2 explicite à quelle approche spécifique ont été attribués dans ce travail les différents gains identifiés.

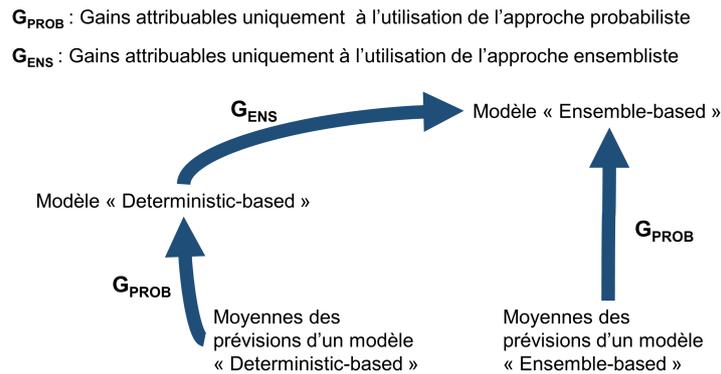


Figure 3.8.2: Gains des approches probabilistes et ensemblistes.

Dans cette partie, nous mettons en lumière les gains obtenus par l'approche probabiliste pour les différentes méthodes de prévision considérées. Le MC-CRPS permet justement l'estimation de ces gains probabilistes effectifs G_e en pourcentage, à la façon d'un skill score, suivant :

$$G_e = 1 - \frac{CRPS}{CRPS_{m1}} = \frac{G_2 + G_+}{CRPS_{m1}}. \quad (3.16)$$

Les résultats obtenus sont présentés sur la table 3.8.1. Ces gains effectifs peuvent être comparés aux gains potentiels déterminés dans le chapitre 1.

Modèle	HAW	DR	SP	PAL	TIR	LAN
AnEn _c	29.5	21.1	28.0	27.6	27.6	27.5
LQR _c	30.5	20.6	30.5	28.9	26.9	29.0
AnEn _m	30.2	21.6	29.7	28.3	27.0	26.7
LQR _m	31.4	21.9	31.9	29.0	27.3	28.0
raw Ensemble	16.1	19.0	21.7	25.4	22.1	24.1
LQR _s	31.6	27.4	34.6	32.0	28.0	31.1
t_NGR	30.5	26.4	30.4	30.4	28.2	28.6
NR GEV	31.7	26.8	34.5	29.5	27.8	28.2

Table 3.8.1: Gains effectifs de qualité de l'approche probabiliste (en %) pour les 6 sites considérés. Gris : approche "deterministic-based", cyan : méthode utilisant la moyenne des membres (approche "ensemble-based"), vert : méthode utilisant la moyenne des membres et d'autres informations de l'ensemble (approche "ensemble-based"). Les valeurs en gras indiquent les meilleurs gains pour un site donné.

Les gains effectifs en qualité obtenus dépendent du site et de la méthode de prévision. En règle générale, les méthodes utilisant l'écart-type des prévisions d'ensemble obtiennent des gains effectifs plus importants, ce qui confirme l'intérêt de l'utilisation des prévisions d'ensemble. On peut noter que les gains effectifs maximums sont très proches des gains théoriques maximums calculés dans le chapitre 1 pour des prévisions non biaisées (de l'ordre de 30 %).

Chapitre 4

Valeur des prévisions probabilistes et liens avec la qualité

Résumé

Ce chapitre traite de l'utilisation pratique des prévisions solaires. Nous recherchons comment quantifier ou mesurer les bénéfices apportés par une prévision dans un contexte d'utilisation opérationnelle. Cette propriété, appelée valeur d'une prévision, reste moins étudiée que la qualité. C'est pourquoi nous cherchons aussi à établir des liens entre la qualité définie et étudiée dans le chapitre 1 et la valeur d'une prévision. Cette valeur dépend cependant de l'utilisation qui est faite de la prévision, et des prises de décision qu'elle aide à optimiser. Comme l'évaluation des gains n'est pas identique pour tous les utilisateurs, il est nécessaire de s'interroger en premier lieu sur la structure des coûts s'appliquant à ces prises de décision. C'est pourquoi les objectifs de ce chapitre sont finalement multiples et consistent à présenter à la fois comment modéliser les coûts associés à une prise de décision à travers des fonctions appelées fonctions pertes et comment cette modélisation affecte la prise de décision finale. Après une brève étude bibliographique qui s'arrête sur la valeur des prévisions de variables binaires, les recherches s'orientent vers les variables continues et aboutissent à la proposition d'une méthodologie permettant d'évaluer la valeur d'un modèle de prévision *a priori*, en connaissant la qualité du modèle et sous certaines hypothèses.

Pour illustrer concrètement les concepts abordés dans cette étude théorique, l'exemple de la valorisation économique de la production électrique d'un parc photovoltaïque est utilisé. Les données de production d'un parc de production solaire photovoltaïque simulé sont utilisées pour modéliser la valorisation économique de la production sur les marchés J+1 de l'électricité. Une grande attention est portée sur les bourses de l'électricité et leur fonctionnement, afin d'en comprendre les contextes réglementaires et les enjeux. On met en évidence le caractère variable des prix de l'électricité, qui nécessitent une prévision déterministe dédiée. Pour cet exemple, les valeurs des différents modèles sont alors quantifiées par le bénéfice économique apporté à l'utilisateur. Les différents modèles de prévision sont comparés, et les gains des prévisions probabilistes sont mis en évidence. La méthodologie exposée au cours de ce chapitre est utilisée pour expliquer ces résultats, et notamment l'écart de valeur entre les prévisions déterministes et probabilistes. L'importance de la qualité de la prévision des prix sur le résultat final est également soulignée.

Sommaire

4.1	Introduction	125
4.2	Valeur des prévisions de variables binaires	126
4.3	Une méthodologie pour l'évaluation de la valeur des prévisions de variables continues	128
4.3.1	Modèles de prévision utilisés	129
4.3.2	Les fonctions pertes : généralités	130
4.3.3	Les fonctions pertes de type "market-based"	132
4.3.4	Liens avec le <i>quantile score</i>	134
4.3.5	Outils graphiques pour l'évaluation de la valeur	134
4.3.6	Liens avec d'autres métriques de qualité	138
4.3.7	Liens avec le cadre d'évaluation des prévisions de variables binaires	141
4.4	Application à un cas pratique : Valorisation économique d'un parc de production solaire	141
4.4.1	Valorisation financière de la production d'électricité	142
4.4.2	Différentes stratégies d'optimisation	148
4.4.3	Prévision déterministe des prix de déséquilibre	152
4.4.4	Résultats	156
4.5	Conclusions	172
	Annexes	174
4.A	La méthode "Variance Deficit"	174
4.B	Modélisation de la production photovoltaïque	175
4.B.1	Données et présentation du parc de production	176
4.B.2	Suréchantillonnage des prévisions	177
4.B.3	Modélisation de la production électrique	179
4.C	Graphique de <i>quantile score</i> d'une prévision déterministe	180
4.D	Mise en place du modèle de prévision des prix ARIMA	180
4.D.1	Étude des autocorrélations temporelles	180
4.D.2	Variables temporelles et ordres du modèle	182

4.1 Introduction

La grande majorité des travaux de recherche dans le domaine de la prévision solaire est dédiée à l'obtention de prévisions de bonne qualité (Behera et al., 2018). La qualité est aussi souvent utilisée comme critère principal pour discriminer différentes prévisions (Pelland et al., 2013). Pourtant, un reproche souvent formulé de la part du monde industriel vers les travaux académiques est qu'il n'y a *a priori* aucune raison pour que la prévision ayant la meilleure qualité soit la plus intéressante pour l'utilisateur final. Même si tel est le cas, il peut exister différents critères d'évaluation de la qualité, et le choix du bon critère de jugement doit dépendre de l'utilisation faite de la prévision. Par exemple, dans le cas des prévisions déterministes, deux métriques différentes comme le *MAE* et le *RMSE* peuvent classer différemment 2 modèles de prévision. Il est important de comprendre quelle sera la structure des coûts correspondant aux prises de décision de l'utilisateur final avant de choisir quelle métrique utiliser pour évaluer correctement une prévision. Or, c'est bien pour les utilisateurs que sont conçues les prévisions.

Les prévisions probabilistes sont justement conçues pour fournir une information plus complète à l'utilisateur, dans le but que ce supplément d'information lui apporte des bénéfices. Or, l'apparente complexité de cette forme de prévision freine bien souvent les utilisateurs potentiels. Comme, en plus, la mesure de la qualité des prévisions par des scores propres comme le *CRPS* ne reflète pas nécessairement les bénéfices obtenus par l'utilisateur, il est difficile d'expliquer simplement l'intérêt de ce type de prévision. C'est pour ces raisons que la communauté scientifique recherche maintenant des moyens de mesurer les bénéfices retirés par l'utilisation d'une prévision. C'est l'étude de la valeur d'une prévision. Cet attribut a d'abord été mis en lumière par Murphy (1993). Il distingue 3 caractéristiques pouvant être attribuées à une prévision météorologique : la qualité, la valeur et la consistance, mais c'est bien la valeur des prévisions qui intéresse le plus les acteurs des marchés de l'électricité, qu'ils soient des producteurs, des fournisseurs, des courtiers de l'énergie ou des gestionnaires de réseau. Le but de ce chapitre est donc de proposer une méthodologie d'évaluation de la valeur, puis de l'éprouver sur un exemple simple d'utilisation pratique de prévisions probabilistes et déterministes. Il nous semble que l'étude et l'explication de la valeur des prévisions probabilistes seraient à même de créer une passerelle entre le monde académique et le monde industriel, facilitant ainsi la dissémination de cette forme de prévision. Mettre en évidence la valeur ajoutée des prévisions probabilistes par rapport aux prévisions déterministes sur des cas pratiques peut être un moyen de convaincre les utilisateurs de l'intérêt des prévisions probabilistes.

Le présent chapitre aura donc deux buts principaux : d'une part, étudier la valeur des prévisions probabilistes déjà présentées dans les chapitres précédents ; d'autre part, rechercher et souligner les liens éventuels entre qualité et valeur des prévisions. Le bénéfice qu'un utilisateur retire d'une prévision, donc la valeur de celle-ci, peut être de différentes natures : financière, environnementale, technique, logistique, etc. La valeur est donc par définition variable et dépend de ce qui en est attendu de la prévision. Il y a ainsi une multitude de définitions possibles, et il n'est pas possible de définir une unique méthodologie pour l'étudier. Souvent, les études mettent l'accent sur les bénéfices économiques, car ils sont plus simplement comparables entre eux. Les travaux qui se consacrent à l'étude de la valeur des prévisions probabilistes mesurent en général le bénéfice en s'appuyant sur une application réelle, que ce soit dans le cas de prévisions déterministes (Jordan et Piwko, 2011 ; Wang et al., 2016), ou probabilistes (Roulston et al., 2003). Comme la définition

de la valeur dépend de l'utilisateur, il n'est possible d'établir des liens entre valeurs et qualité que dans des cas particuliers. C'est néanmoins l'approche que nous allons suivre dans ce chapitre.

La vente sur le marché de l'électricité J+1 de la production photovoltaïque d'un parc solaire situé sur le site de l'aéroport d'Adélaïde (Australie) a été choisie comme cas d'étude pour ce chapitre. Le parc solaire est présenté dans l'Annexe 4.B, ainsi que tous les détails concernant les méthodes de modélisation de la production solaire en fonction de l'ensoleillement. Les modèles de prévision sont eux détaillés dans la partie 4.3.1. Ces prévisions sont ensuite utilisées dans les sections 4.3 et 4.4. La section 4.3 est consacrée à une recherche théorique sur la valeur des prévisions dont l'utilisation est soumise à une structure de coût particulière couramment rencontrée par les acteurs des marchés de l'énergie. Cette recherche aboutit à la proposition d'une méthodologie d'évaluation de la valeur d'un système de prévision. Cette méthodologie nécessite l'utilisation de 2 outils graphiques : Le PEVC ("Potential Economic Value for Continuous variables") et le graphique de répartition des risques. Utilisés ensemble, ils forment un nouvel outil appelé le EVC ("Effective Value for Continuous variables") et permettent de visualiser la valeur d'un système de prévision.

La partie 4.4 présente ensuite une application de cette méthodologie. Elle étudie la valeur des prévisions sur un cas pratique qui consiste à valoriser la production électrique par une vente sur les marchés de l'électricité. Cela permet de disposer d'une métrique simple permettant l'évaluation de la valeur : le bénéfice retiré de la vente pendant une période de test. Cette vente est étudiée dans 3 cas différents : une vente sur le marché français de l'électricité, une vente sur le marché portugais et une vente sur le marché suisse. Pour ces cas d'étude, le graphique EVC est utilisé comme outil d'explication des résultats obtenus.

4.2 Valeur des prévisions de variables binaires

Les études sur la valeur des prévisions probabilistes se sont pour l'instant majoritairement attachées aux prévisions de variables binaires (Thornes et Stephenson, 2001 ; Zhu et al., 2002 ; Wilks, 2001). Cette partie brosse un bref panorama des méthodes d'évaluation de la valeur de telles prévisions.

Considérons un risque météorologique qui peut être modélisé comme suit : un événement, cible d'une prévision probabiliste, peut se produire ou non. Si tel est le cas, la réalisation de cet événement induit pour un acteur industriel une perte P (qui peut être économique ou d'une autre nature). Cet acteur peut choisir de se protéger, ou non, contre ce risque, moyennant un coût C . Cette situation est appelée le "dilemme coût-perte" ("Cost-Loss situation"). Classiquement, les coûts subis par l'acteur à l'issue d'une expérience se résument par la table de contingence 4.2.1

		Observation	
		Oui	Non
Protection	Oui	Succès C	Fausse alarme C
	Non	Échec P	Rejet correct 0

TABLEAU 4.2.1 : Table de contingence des issues possibles et des coûts associés après une prise de décision et une observation d'un évènement binaire.

Murphy (1966) montre que le comportement optimal consiste à se protéger dès lors que la probabilité que l'évènement survienne est supérieure au ratio Coût-Perte ("Cost-Loss ratio") C/P . On peut alors associer à une prévision probabiliste binaire le coût moyen subi par un utilisateur s'aidant de celle-ci pour prendre des décisions. Pour un ratio fixé, ce coût moyen Em est calculé par :

$$Em = (a + b) \times C + c \times P, \quad (4.1)$$

où a , b et c sont les fréquences relatives respectives des succès, des fausses alarmes et des échecs. Il est alors naturel de tracer la "courbe des coûts" d'une prévision qui représente Em en fonction de $x=C/P$, variant entre 0 et 1. On peut alors démontrer (Wilks, 2014) que l'aire sous cette courbe est liée au *Brier score* de la prévision par :

$$BS = \int_0^1 Em(x)dx - Em_{parfait}, \quad (4.2)$$

où $Em_{parfait}$ est le coût moyen de la prévision parfaite, qui est définie par des probabilités nulles de fausse alarme et d'échec. Autrement dit, le *Brier score*, qui est un outil de mesure de la qualité d'une prévision binaire (cf. chapitre 1) permet aussi de quantifier la valeur d'une telle prévision probabiliste, à condition que la distribution des ratios C/P soit uniforme sur $[0; 1]$. Plus cette distribution s'éloigne de la distribution uniforme, moins le *Brier score* est un indicateur fiable de la valeur.

Comme un modèle simple de prise de décision basé sur le ratio climatologique est applicable sans implémenter de modèle de prévision spécifique, le coût moyen d'un modèle est souvent comparé à celui de la climatologie. Buizza (2008) calcule cette comparaison à la manière d'un "skill score" et définit ce faisant la valeur économique relative (ou "REV") d'un système de prévision :

$$REV = \frac{Em - Em_{CLIM}}{Em_{parfait} - Em_{CLIM}} \quad (4.3)$$

où Em_{CLIM} est le coût moyen issu de la prévision climatologique. Il est courant de tracer le graphique représentant la valeur économique relative en fonction du ratio C/P . Il s'agit d'un outil d'évaluation de la valeur d'une prévision binaire (Richardson, 2000). Classiquement, seules les valeurs de REV supérieures à 0 sont tracées. L'exemple d'un tel graphique, appelé graphique de valeur économique potentielle ("PEV"), est présenté sur la figure 4.2.1.

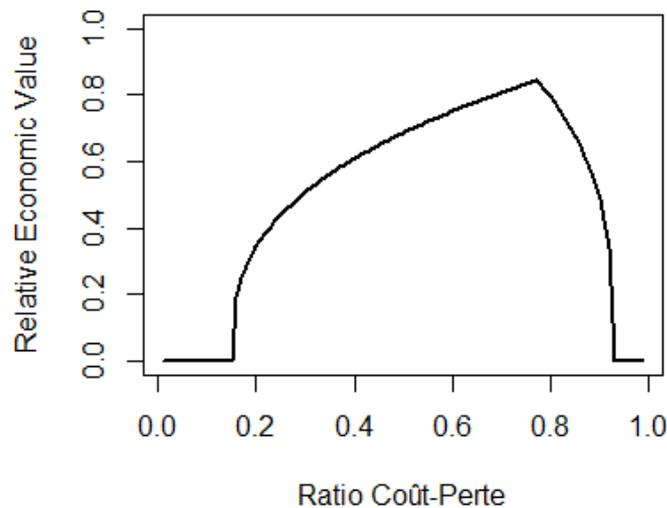


FIGURE 4.2.1 : Exemple d'un graphique de PEV.

La figure 4.2.1 montre que la prévision évaluée apporte un gain de valeur par rapport à la climatologie pour tous les ratios entre 0.17 et 0.92. Sa valeur maximale est pour un ratio de 0.78, avec une valeur économique relative de 0.82.

4.3 Une méthodologie pour l'évaluation de la valeur des prévisions de variables continues

Dans cette partie, la valeur des prévisions est abordée dans un cadre très général. Les résultats obtenus peuvent donc s'appliquer à des prévisions solaires, mais aussi à des prévisions probabilistes générées dans le cadre d'autres applications (banque, assurance etc.). L'objectif de cette partie est de conduire une investigation sur les liens qui peuvent être établis entre qualité et valeur d'une prévision. En particulier, on recherche si des outils de diagnostic utilisés pour mesurer la qualité des prévisions peuvent aussi être pertinents pour juger *a priori* de la valeur de la prévision.

Dans toute cette partie, on considère la prévision probabiliste d'une variable aléatoire X . Cette prévision n'a de valeur que dans un contexte de prise de décision, où l'utilisateur se sert de la prévision pour prendre une décision quantitative V (quel montant investir, quelles mesures de protection prendre face à un risque, etc.). Dans le cas des prévisions de production photovoltaïque utilisées par un acteur qui souhaite valoriser sa production sur les marchés de l'électricité, X est la production électrique et la décision V à prendre porte généralement sur le volume d'électricité à vendre. Il apparaît nécessaire de traiter la problématique de la mise en place d'un processus d'optimisation (ou d'aide à la prise de décision) associée à une prévision, ce qu'on effectue dans la partie 4.3.2. Le rôle prépondérant de la modélisation des coûts associés à la prise de décision est mis en évidence. La question de savoir si des métriques d'évaluation de la qualité peuvent correspondre à la structure des coûts est étudiée à l'aide d'une recherche théorique (partie 4.3.3). Cette

recherche montre un lien fort entre la valeur et le “*quantile score*” (cf. partie 4.3.4) et aboutit finalement à la proposition d’une méthodologie pour l’évaluation de la valeur d’un système de prévision. Elle suggère l’utilisation de deux graphiques complémentaires présentés dans la partie 4.3.5 : le PEVC (“Potential Economic Value for Continuous variables”) et le “graphique de répartition des risques”. Ces deux graphiques combinés forment le graphique EVC (“Effective Value for Continuous Variable”). Une illustration de l’utilisation du graphique EVC est proposée, à travers l’étude de la différence de valeur des modèles probabilistes et déterministes du tableau 4.3.1. Finalement, dans la partie 4.3.6, une investigation est conduite sur les liens entre la valeur et deux autres scores utilisés pour la mesure de la qualité des prévisions : l’*ignorance score* et le *CRPS*.

4.3.1 Modèles de prévision utilisés

Pour cette étude, les prévisions d’ensemble brutes ECMWF ont été utilisées comme prévisions d’éclairage énergétique. Elles ont été transformées en prévision de production d’électricité d’un parc de production photovoltaïque de 1 MW crête au pas horaire. Le détail des étapes permettant de fournir une prévision de production à partir de la prévision du GHI est explicité dans l’Annexe 4.B. Ces prévisions de production étant calculées à partir d’un ensemble brut, il est nécessaire de les calibrer à l’aide de techniques de post-traitement. Des modèles de prévision incluant des techniques de calibration de la prévision brute ECMWF déjà présentées dans le chapitre 1 ont été utilisés. Le tableau 4.3.1 résume ces différentes méthodes :

Modèle	Type	Variables utilisées	Nom
Contrôle	Déterministe	Prévision de contrôle	control
Analog Ensemble	Probabiliste	Prévision de contrôle	AnEn _c
Régression quantile linéaire	Probabiliste	Prévision de contrôle	LQR _c
Moyenne	Déterministe	Moyenne des membres	mean
Analog Ensemble	Probabiliste	Moyenne des membres	AnEn _m
Régression quantile linéaire	Probabiliste	Moyenne des membres	LQR _m
ensemble brut	Probabiliste	Tous les membres	raw
Variance Deficit	Probabiliste	Tous les membres	VD
Régression quantile linéaire	Probabiliste	Tous les membres	LQR _s
Régression gaussienne tronquée	Probabiliste	Tous les membres	t_NGR
Régression “GEV”	Probabiliste	Tous les membres	NR GEV
Parfait	Déterministe	Observations	parfait

TABLEAU 4.3.1 : Les modèles de prévision utilisés, leur type, leur approche, leurs variables d’entrée et leur couleur associée. La couleur associée à un modèle se réfère aux variables utilisées, dans le but de rester consistant avec le chapitre 3.

Tous ces modèles ont été construits à partir d’une base d’apprentissage d’un an (2016) et sont testés sur une période d’un an (2017). Pour respecter la terminologie définie dans la chapitre 3, les modèles n’utilisant que la prévision de contrôle peuvent être qualifiés de modèles “Deterministic-based” alors que tous les autres sont “Ensemble-based”. Le modèle VD, non présenté dans le chapitre 3, est expliqué dans l’Annexe 4.A. Le modèle parfait n’est pas un modèle de prévision. Il correspond à une prévision déterministe sans erreur, donc égale aux observations.

4.3.2 Les fonctions pertes : généralités

Les coûts induits par le processus de prise de décision peuvent être modélisés par une fonction qui dépend de la différence D entre la quantité observée O et la décision prise V .

$$D = O - V. \quad (4.4)$$

On appelle cette fonction la “fonction de perte” (ou fonction $L(D)$, ou encore $L(V, O)$). Cette fonction peut refléter uniquement des coûts financiers, humains, techniques, logistiques, ou être un mélange de ces différents coûts (Holttinen, 2005). Elle peut être théoriquement de n’importe quelle nature, mais est minimale pour $D = 0$. Cela traduit simplement le fait qu’une prévision sans erreur minimise les coûts pour l’utilisateur. Trois exemples possibles de fonctions pertes sont présentés sur la figure 4.3.1.

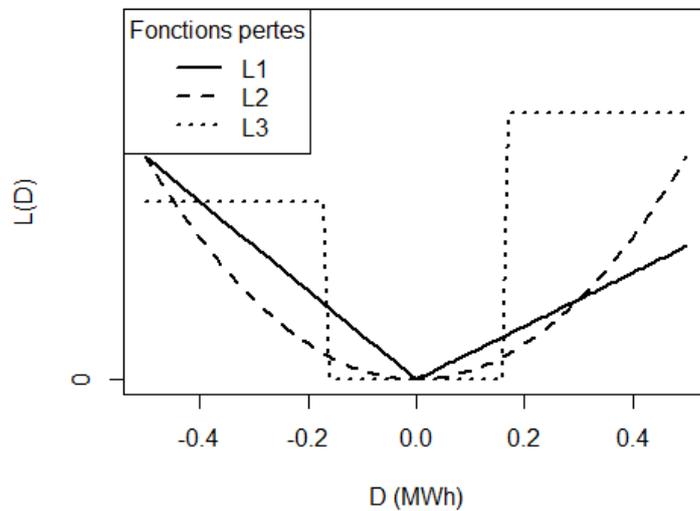


FIGURE 4.3.1 : Trois exemples de fonctions pertes.

À l’aide de la fonction perte, on connaît les coûts associés à une prise de décision V et une observation O . Considérant que la production d’électricité est une variable aléatoire notée X de fonction PDF q , on peut définir pour chaque décision possible V , l’espérance des coûts associés $\mathcal{P}(V) = \mathbb{E}[L(V, X)]$ que l’on notera dans la suite $\mathcal{P}(V) = \mathbb{E}_q[L]$. Dans ce formalisme, la décision optimale du point de vue de l’utilisateur est celle qui minimise $\mathcal{P}(V)$. Une prévision \hat{f} de la densité de probabilité réelle q permet de calculer une approximation $\hat{\mathcal{P}}(V) = \mathbb{E}_{\hat{f}}[L]$ de $\mathcal{P}(V)$ (Bourry et al., 2008). La minimisation de cette fonction $\hat{\mathcal{P}}(V)$ permet alors d’optimiser la prise de décision V .

Le théorème de transfert indique que, pour une variable aléatoire X de densité f et toute fonction ϕ , on a l’identité :

$$\mathbb{E}[\phi(X)] = \int_{\mathbb{R}} \phi(x) f(x) dx. \quad (4.5)$$

L’application de ce théorème permet de calculer $\hat{\mathcal{P}}(V)$:

$$\hat{\mathcal{P}}(V) = \mathbb{E}_{\hat{f}}[L(V)] = \int_{\mathbb{R}} L(V, x) \hat{f}(x) dx. \quad (4.6)$$

Optimisation avec une prévision déterministe

Une prévision déterministe est constituée d'une seule valeur possible de production photovoltaïque \hat{F}_d . Dans ce cas, on peut tracer la fonction perte L_d "associée à cette prévision" (c'est-à-dire centrée telle que : $D = 0$ pour $V = \hat{F}_d$) car l'équation 4.6 se réduit à :

$$\hat{\mathcal{P}}(V) = L_d. \quad (4.7)$$

La décision optimale compte-tenu des informations disponibles est alors nécessairement égale à \hat{F}_d , la valeur donnée par la prévision. Par ailleurs, la fonction perte étant nulle en $D = 0$, la prévision déterministe amène à considérer que le coût induit par cette décision est nul.

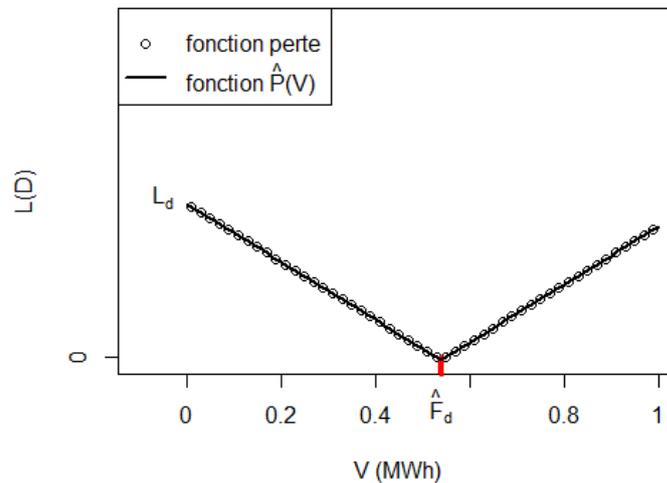


FIGURE 4.3.2 : Exemple d'une optimisation de minimisation des pertes pour une prévision déterministe (\hat{F}_d). La vente qui minimise l'espérance des pénalités, mise en évidence par un segment rouge, est celle qui correspond au minimum de la fonction $\hat{\mathcal{P}}$, égale à L_d . Cette vente est donc égale à \hat{F}_d .

Optimisation avec une prévision d'ensemble

Plusieurs manières d'effectuer une optimisation à partir d'une prévision probabiliste peuvent être considérées (Pinson et al., 2007). La méthode suivie dans ce travail est résumée ci-dessous.

Une prévision d'ensemble peut être vue comme un ensemble de prévisions déterministes (\hat{F}_i), toutes équiprobables. Quelle que soit la fonction perte, il est possible d'utiliser simplement cette forme de prévision pour calculer numériquement $\hat{\mathcal{P}}$, et ainsi optimiser la prise de décision V . Une fonction perte L_{E_i} est d'abord associée à chacun des membres (E_i^p) de la prévision d'ensemble. Comme la prévision d'ensemble fournit un ensemble de scénarios équiprobables, la moyenne de toutes ces fonctions pertes est une prévision $\hat{\mathcal{P}}(V)$ de $\mathcal{P}(V)$ et son minimum indique la décision optimale. Mathématiquement, cela se traduit par l'utilisation d'une version discrète du théorème de transfert, en considérant que

$\hat{f}(x) = \frac{1}{M} \sum_{i=0}^M \delta(x - E_i^p)$, où δ est la distribution de Dirac :

$$\hat{\mathcal{P}}(V) = \int_0^\infty L(V, x) \hat{f}(x) dx = \frac{1}{M} \sum_{i=0}^M L(V, E_i^p). \quad (4.8)$$

Cette méthode est illustrée sur la figure 4.3.3.

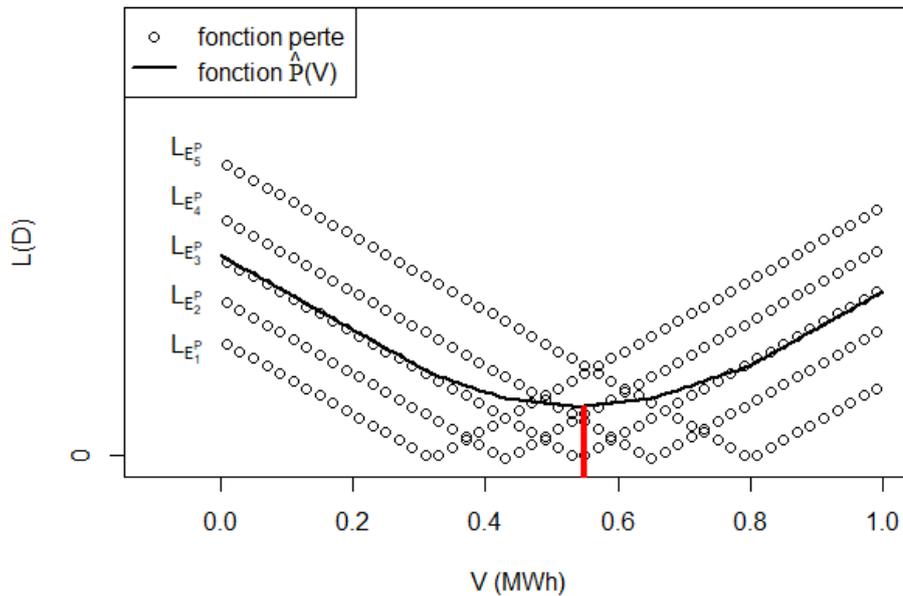


FIGURE 4.3.3 : Exemple d’une optimisation de minimisation des pertes pour un ensemble de 5 membres (M_1, \dots, M_5). La vente qui minimise l’espérance des pénalités est celle qui correspond au minimum de $\hat{\mathcal{P}}$ (la moyenne des fonctions pertes de tous les membres), indiqué par le segment rouge épais. La valeur de cette fonction en ce point indique le coût attendu pour cette décision.

Par définition, la fonction $\hat{\mathcal{P}}$ donne en tout point une estimation probabiliste du coût subi par l’utilisateur. Ainsi, en plus de donner la décision optimale ($V = \arg \min \hat{\mathcal{P}}$), cette fonction peut être utilisée par l’utilisateur pour estimer son coût attendu. C’est la valeur minimale de la fonction $\hat{\mathcal{P}}$.

4.3.3 Les fonctions pertes de type “market-based”

Le cas où la fonction perte est affine par morceaux, avec deux pentes différentes (une pour $D < 0$ et une autre pour $D > 0$) et un minimum nul en $D = 0$ est couramment étudié dans la littérature scientifique consacrée à l’achat/vente d’énergie éolienne (Linnet, 2005 ; Bremnes, 2004), et peut être une bonne approximation des coûts subis par un utilisateur dans le contexte de l’optimisation de la vente d’énergie éolienne ou solaire (Pinson et al., 2004). Elle correspond en effet aux coûts financiers directs d’un utilisateur qui vend son énergie sur les marchés de l’électricité. Ce type de fonction perte sera repris dans

l'application de la section 4.4. La fonction L1 du graphique 4.3.1 correspond à cette définition. Suivant la terminologie proposée par Pinson et al. (2007), on appellera des fonctions de cette forme des fonctions pertes “market-based”. Dans cette partie, on étudie uniquement des fonctions pertes de cette forme. On appellera P_- la valeur absolue de la pente de cette fonction pour $D < 0$, P_+ la valeur absolue de cette pente pour $D > 0$, R le ratio $\frac{P_-}{P_+ + P_-}$ caractérisant la dissymétrie entre les 2 pentes et γ la somme $P_+ + P_-$, qui peut être vue comme une mesure du risque auquel est exposé l'utilisateur.

Dans le cas particulier des fonctions pertes de type “market-based”, il est possible de calculer analytiquement la quantité qui minimise l'espérance des coûts pour l'utilisateur, à partir d'une prévision probabiliste sous la forme d'une fonction de répartition (ou fonction “CDF”). Supposons que la prévision probabiliste de la variable aléatoire X soit sous la forme d'une fonction PDF $\hat{f}(x)$ associée à la CDF \hat{F} , qui est une image de la fonction PDF réelle et inconnue q (de CDF associée \mathcal{Q}). L'optimisation des coûts de l'utilisateur peut s'exprimer par :

$$J = \arg \min_V \left[\hat{\mathcal{P}}(V) \right]. \quad (4.9)$$

Or, dans le cas de fonctions pertes de type “market-based”, on peut écrire ¹ :

$$\hat{\mathcal{P}}(V) = \left(\int_{x=0}^{x=V} (P_- \times (V - x)) \hat{f}(x) dx + \int_{x=V}^{x=+\infty} (P_+ \times (x - V)) \hat{f}(x) dx \right), \quad (4.10)$$

$$\hat{\mathcal{P}}(V) = (P_+ + P_-) \times \left[\frac{P_-}{P_+ + P_-} \int_{x=0}^{x=V} (V - x) \hat{f}(x) dx + \left(1 - \frac{P_-}{P_+ + P_-} \right) \int_{x=V}^{x=\infty} (x - V) \hat{f}(x) dx \right]. \quad (4.11)$$

On remarque que les quantités $\gamma = (P_+ + P_-)$ et $R = \frac{P_-}{P_+ + P_-}$ ont une importance prépondérante dans le calcul de J . On peut réécrire $\hat{\mathcal{P}}(V)$ sous la forme suivante :

$$\hat{\mathcal{P}}(V) = \gamma \times \left[(1 - R) \int_{x=0}^{x=V} (V - x) \hat{f}(x) dx + R \int_{x=V}^{x=\infty} (x - V) \hat{f}(x) dx \right]. \quad (4.12)$$

On peut reconnaître l'expression de l'espérance de la fonction de perte Pinball “ $\rho_{\tau=R}$ ” pour une prévision égale à V (cf. partie 1.A). Ainsi on trouve :

$$\hat{\mathcal{P}}(V) = \gamma \times \mathbb{E} \left[\rho_{\tau=R}(V) \right], \quad (4.13)$$

où ρ_τ est la fonction de perte Pinball associée au niveau de probabilité τ . On sait déjà que l'espérance de cette fonction est minimale pour une prévision égale à $\hat{F}^{-1}(R)$ (cf. chapitre 1), et on peut déduire de cette optimisation analytique que $\hat{F}^{-1}(R)$ est la décision V optimale. Il est à noter que ce résultat concorde avec les conclusions de Linnet (2005). Cette résolution fait apparaître les rôles respectifs de R , qui représente la dissymétrie entre les coûts associés aux erreurs de prévision négatives et positives, et de γ , qui est de son côté une mesure du niveau de risque associé à une prise de décision.

¹La vente d'électricité ne peut prendre que des valeurs positives, ce qui permet de fixer pour ce cas particulier la borne inférieure de l'intégrale à 0.

4.3.4 Liens avec le *quantile score*

Le lien entre la valeur d'une prévision dans le contexte de coûts de type "market-based" et la fonction ρ est matérialisé par l'équation 4.13. Il peut aussi être mis en évidence par les coûts finaux subis par l'utilisateur suite à l'optimisation. Ces derniers sont donnés par la fonction perte $L(D)$ et sont égaux à :

$$\begin{cases} \gamma \times R \times D & \text{si } D < 0, \\ \gamma \times (R - 1) \times D & \text{si } D > 0. \end{cases} \quad (4.14)$$

On reconnaît l'expression de la fonction de perte Pinball ρ associée au *quantile score* de niveau de probabilité $\tau = R$, pondéré par la somme γ . On peut donc réécrire ces coûts sous la forme :

$$\gamma \times \rho_{\tau=R}(V, O). \quad (4.15)$$

L'équation 4.15 démontre en particulier que seule la qualité de la prévision du niveau de probabilité $\tau = R$ a de la valeur pour l'utilisateur, alors que la prévision de tous les autres niveaux de probabilité a une valeur nulle. Il n'est donc *a priori* pas nécessaire que l'évaluation de la qualité de la prévision se fasse sur l'ensemble de la CDF pour déterminer sa valeur. En pratique, les utilisations d'un système de prévision incluent de multiples couples prévision/observation. Dans un tel cas, les coûts associés au modèle de prévision peuvent être calculés par la somme des fonctions ρ :

$$\sum_{n=1}^N \rho_{\tau,n}(V_n, O_n), \quad (4.16)$$

où N le nombre de couples prévision/observation. Or, il a été mis en évidence dans le chapitre 1 que :

$$QS_{\tau} = \frac{1}{N} \sum_{n=1}^N \rho_{\tau,n}(\hat{F}_{\tau}, O_n). \quad (4.17)$$

On voit alors apparaître un lien fort avec le *quantile score*, qui se trouve être un outil d'évaluation de la qualité qui rend correctement compte des structures de coûts de type "market-based". À partir de ce lien fort démontré entre *quantile score* et valeur de la prévision dans le cas particulier où la fonction perte est "market-based", nous proposons dans la section 4.3.5 des outils d'évaluation de la valeur des prévisions probabilistes pour les variables continues et des fonctions pertes de type "market-based".

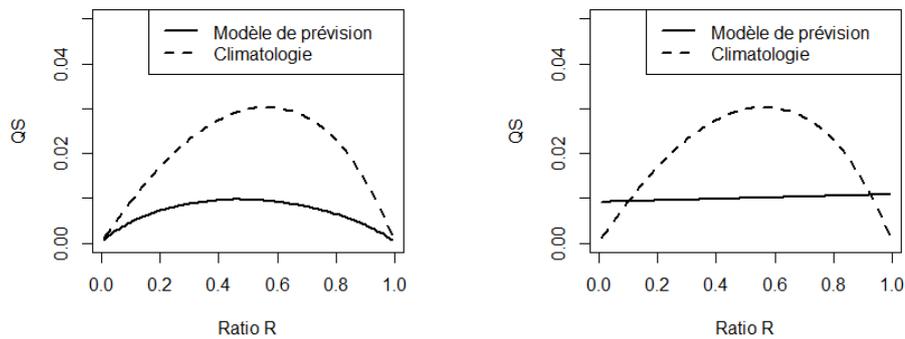
4.3.5 Outils graphiques pour l'évaluation de la valeur

Le graphique de *quantile score*

Alors que pour les prévisions binaires, la variable importante est le "cost-loss ratio" C/P (cf. section 4.2), dans le cas des variables continues, c'est le ratio R qui a été identifié comme variable-clé. Nous proposons donc de calculer le *quantile score* d'une prévision identifiée par sa CDF \hat{F} pour tous les niveaux de probabilité entre 0 et 1, à un pas assez faible (nous avons choisi dans cette étude un pas de 0.01), puis de tracer le *quantile score* de la prévision en fonction du niveau de probabilité τ . Le graphique ainsi produit a déjà été utilisé par Lauret et al. (2019) pour l'évaluation de la qualité des prévisions

probabilistes. Nous proposons ici d'étendre son utilisation à l'évaluation de la valeur de prévisions probabilistes dans le cas de fonctions de perte "market-based". Ce graphique indique la valeur de la prévision pour tout ratio R ; il peut donc être interprété comme une représentation du potentiel de la prévision à générer des coûts moyens faibles pour un utilisateur qui fait face à une structure de type "market-based".

Les *quantiles scores* de deux modèles de prévisions, l'une déterministe (modèle "mean"), l'autre probabiliste (modèle "VD"), sont tracés sur la figure 4.3.4. L'étude de ces deux graphiques permet la comparaison de la valeur de ces deux prévisions pour des fonctions de perte de type "market-based".



(a) *Quantile score* du modèle VD (pré- vision probabiliste). (b) *Quantile score* de la moyenne des membres (prévision déterministe).

FIGURE 4.3.4 : *Quantiles scores* pour les modèles de prévision de la production électrique "VD" et "mean".

On rappelle que le *quantile score* est négativement orienté (une valeur faible dénote une meilleure qualité). On peut constater sur la figure 4.3.4 que le modèle VD domine nettement la climatologie pour tous les ratios R . À l'inverse, le modèle "mean" ne bat pas la climatologie pour les ratios proches de 0 ou 1. De plus, le modèle "VD" a plus de valeur que le modèle "mean" pour tous les ratios, mais cette différence est nettement plus importante pour les ratios extrêmes (proches de 0 et 1) que pour les ratios médians. C'est une différence fondamentale entre un modèle de prévision déterministe et une prévision probabiliste. On peut démontrer que le graphique de *quantile score* d'une prévision déterministe est nécessairement une droite (cf. Annexe 4.C). À cause de cette caractéristique, la prévision déterministe ne peut jamais être performante pour tous les ratios R . En pratique, une prévision déterministe est en général une prévision proche de la médiane. On peut alors s'attendre à ce qu'elle soit très peu performante en comparaison des prévisions probabilistes pour des ratios éloignés du niveau de probabilité $R = 0.5$.

Un outil d'évaluation de la valeur potentielle : le "PEVC"

Cette partie est dédiée à la présentation d'un outil proposant d'évaluer la valeur économique d'une prévision probabiliste d'une variable continue pour les fonctions pertes de type "market-based" (cf. section 4.3.2). Nous proposons ici d'utiliser un graphique construit sur le même principe que le graphique de PEV pour les prévisions binaires (cf.

partie 4.2). Pour construire un cadre d'évaluation cohérent avec celui de l'évaluation de la valeur des prévisions binaires, nous proposons de comparer la valeur des prévisions avec la valeur de la climatologie empirique de la période d'apprentissage (tracée également sur le graphique 4.3.4). Comme recommandé par Ben Bouallegue (2017), le *quantile skill score* de la prévision est calculé pour chaque niveau de probabilité, grâce à :

$$QSS = 1 - \frac{QS}{QS_{clim}}. \quad (4.18)$$

À l'inverse du *quantile score*, un *QSS* plus élevé indique un meilleur modèle de prévision. Une prévision parfaite obtient $QSS = 1$ alors que $QSS < 0$ indique que le modèle testé est moins performant que la prévision climatologique. Nous traçons ce *quantile skill score* pour tous les ratios de l'intervalle $[0, 1]$ et nous proposons d'utiliser ce graphique (que nous appellerons PEVC pour "Potential Economic Value of Continuous variables" par analogie au PEV) comme une mesure de la valeur potentielle des modèles de prévision, par rapport à la climatologie. Dans la figure 4.3.5, seules les valeurs de $QSS > 0$ indiquant une performance meilleure que celle de la climatologie sont représentées.

Sur la figure 4.3.5, les graphiques de PEVC de 3 modèles de prévision utilisés dans cette étude permettent de comprendre les bénéfices des approches probabilistes et ensemblistes.

4.3.1

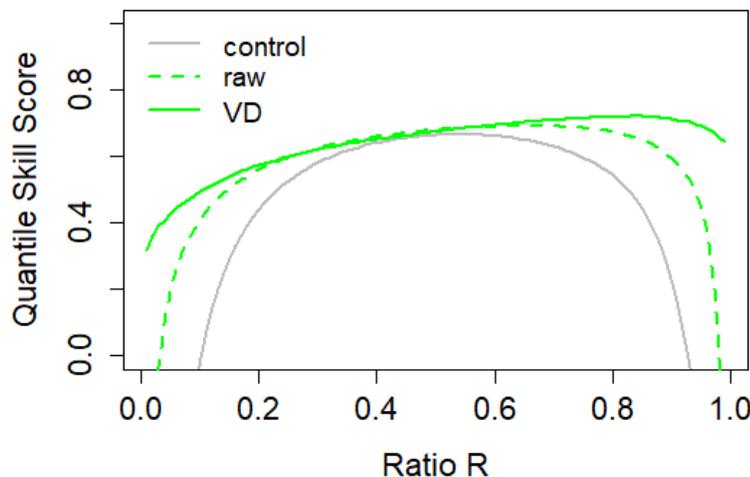


FIGURE 4.3.5 : PEVC de 3 modèles de prévision.

À nouveau, on peut voir que la différence entre les modèles probabilistes et les modèles déterministes est très claire pour les niveaux de probabilité "extrêmes". Pour des cas d'application réels, où la prévision déterministe est non biaisée et indique une prévision proche de la médiane, plus le ratio R est éloigné de $\frac{1}{2}$, plus la différence de valeur entre une prévision probabiliste et une prévision déterministe sera en faveur de la prévision probabiliste. De la même façon, on constate que la calibration de l'ensemble brut améliore principalement la valeur des prévisions pour les ratios R extrêmes.

Cas d'utilisation pratique d'un système de prévision : le “graphique de répartition des risques”

Pour la plupart des cas d'utilisation pratique d'un système de prévision, comme par exemple l'optimisation de la vente d'électricité, R et γ varient et un couple (R_n, γ_n) est associé à chaque prévision, où n varie entre 1 et N , le nombre total de couples prévision/observation. Le ratio R devient une variable dépendante de la structure du marché, de distribution *a priori* quelconque entre 0 et 1. Par conséquent, les coûts finaux pour l'utilisateur peuvent être modélisés correctement par le *quantile score* de niveau de probabilité variable, dont la distribution est égale à la celle du ratio R , pondéré par la somme γ . Ceci explique dans ce formalisme les importances cruciales du ratio R et de γ . Dans le cas d'une utilisation réelle, pour jauger de la valeur d'une prévision, ou pour faire un lien entre qualité et valeur, il apparaît alors nécessaire d'étudier la distribution des ratios R pondérés par la somme des γ . On introduit donc, pour chaque valeur possible du ratio R , la somme des γ , “ S_{γ_R} ” définie par :

$$S_{\gamma_R} = \sum_{n=1}^N \gamma_n \times \mathbb{1}_{R_n=R}. \quad (4.19)$$

Nous appellerons dans cette étude une représentation graphique de cette distribution un “graphique de répartition des risques” (cf. figure 4.3.6). Ce graphique montre l'importance relative des pénalités en fonction du niveau de probabilité R dans un cas d'application pratique.

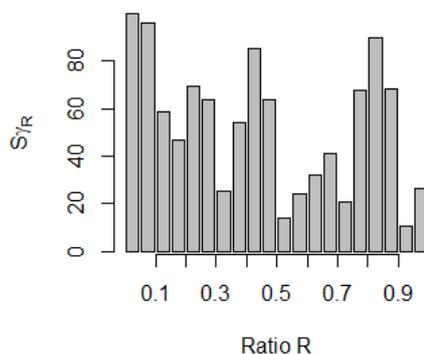


FIGURE 4.3.6 : Exemple d'un graphique de répartition des risques.

Pour l'étude de la valeur d'un système de prévision dans un cas d'application réelle, nous proposons d'associer ce graphique avec le PEVC pour l'analyse de la valeur des prévisions. Ces deux outils sont complémentaires : le PEVC indique la valeur potentielle du système de prévision pour chaque ratio R , quand le graphique de répartition des risques indique l'exposition au risque associée à chaque ratio R pour l'application considérée, c'est-à-dire l'importance relative de chacun des ratios. Autrement dit, le PEVC dépend du modèle de prévision, tandis que le graphique de répartition des risques ne dépend pas de la prévision, et représente la structure de coûts associés au cas d'application. L'étude combinée de ces deux graphiques permet ainsi l'évaluation de la valeur du système de prévision pour

l'application considérée, et dans cette étude nous appelons "EVC" un graphique associant par superposition le PEV et le graphique de répartition des risques. Un exemple simple d'EVC est proposé sur la figure 4.3.7.

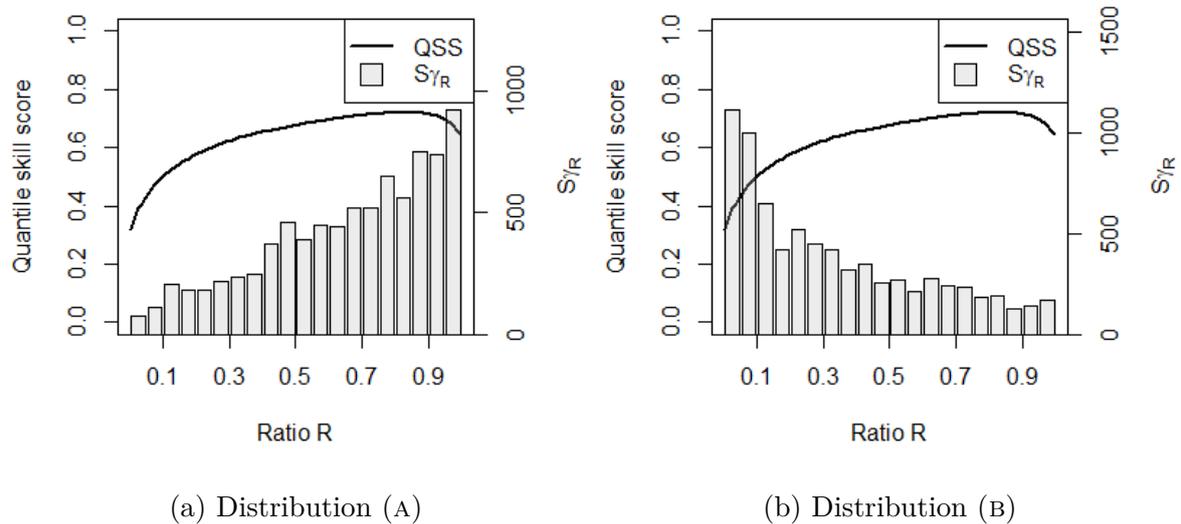


FIGURE 4.3.7 : Graphiques EVC du modèle VD pour deux répartitions des risques S_{γ_R} différentes.

La figure 4.3.7 montre l'importance de la répartition des risques sur la valeur d'un modèle de prévision. Le modèle de prévision VD est plus performant pour les ratios élevés. Ce modèle aura alors une valeur plus élevée pour la distribution A, qui elle-même accorde plus d'importance aux ratios élevés. A l'inverse, le modèle VD n'est pas adapté à la structure de coûts du cas B, car le risque le plus fort est associé aux ratios faibles pour lesquels la prévision est peu performante. Sur cet exemple simple, l'outil graphique EVC montre qu'un modèle de prévision peut avoir une valeur variable selon la distribution des risques associée à son utilisation.

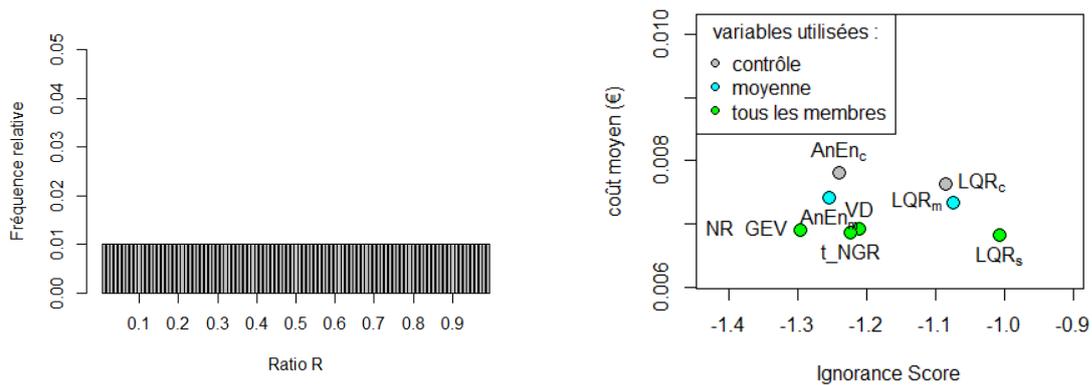
Il est important de noter que cette méthodologie suppose que la distribution γ ne dépend pas de facteurs temporels, en particulier horaires ou saisonniers, et qu'il n'y a pas de corrélation temporelle entre erreur de prévision et structure de coût.

4.3.6 Liens avec d'autres métriques de qualité

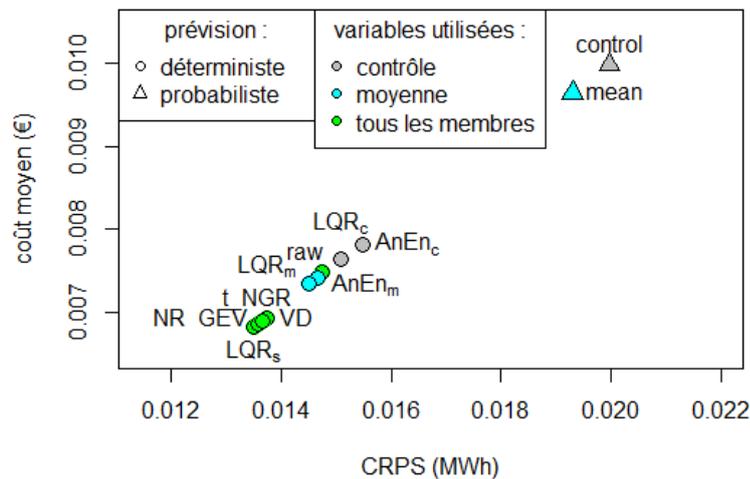
Un lien clair a été établi entre la valeur d'une prévision soumise à une structure de coût de type "market-based" et le *quantile score*. Le lien avec d'autres scores n'est pas aussi évident. Dans cette partie nous étudions la corrélation entre la valeur des prévisions et d'autres métriques permettant d'évaluer la qualité : le *CRPS* et l'*ignorance score*.

Comme vu dans la partie 4.3, la valeur est fortement dépendante du graphique de répartition des risques. Nous étudions donc, pour plusieurs scénarios de distribution des risques arbitrairement choisis, la corrélation entre ces scores et le coût moyen associé à l'utilisation des prévisions, qui est calculé comme le rapport entre le coût total associé à l'utilisation d'un modèle et le nombre de pas de temps de la période de test. 2 distributions différentes sont sélectionnées (voir figures 4.3.8 et 4.3.9 : une distribution dite "uniforme"

où le ratio R est réparti uniformément sur l'intervalle $[0, 1]$ et une distribution "centrée" où R est constant et toujours égal à $\frac{1}{2}$. Nous choisissons de poser $\gamma = 1$ pour chaque couple prévision/observation, de telle sorte que le graphique de répartition de R et le graphique de répartition des risques sont exactement les mêmes. Les graphiques de corrélation de l'ignorance score ne représentent pas les modèles déterministes, car les scores de tels modèles sont infinis (cf chapitre 1). Les résultats obtenus sont présentés sur les figures 4.3.8 et 4.3.9.

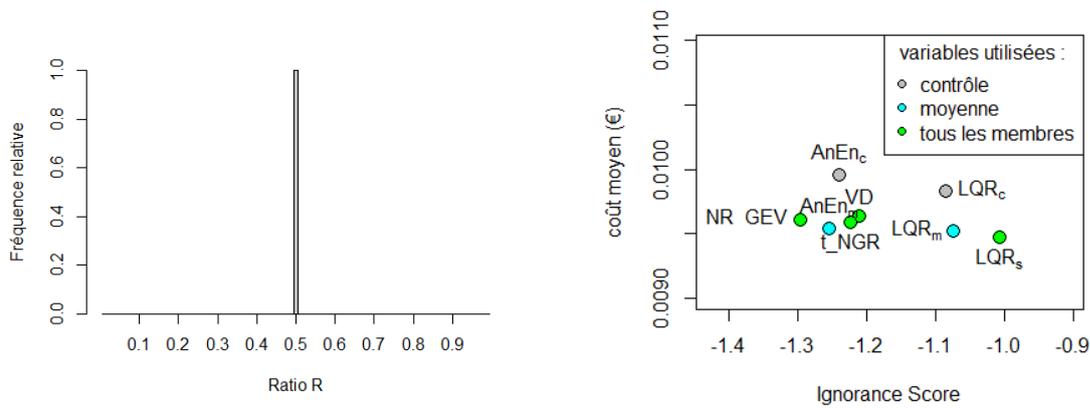


(a) Illustration de la distribution uniforme. (b) Corrélation entre le coût moyen et l'ignorance score.

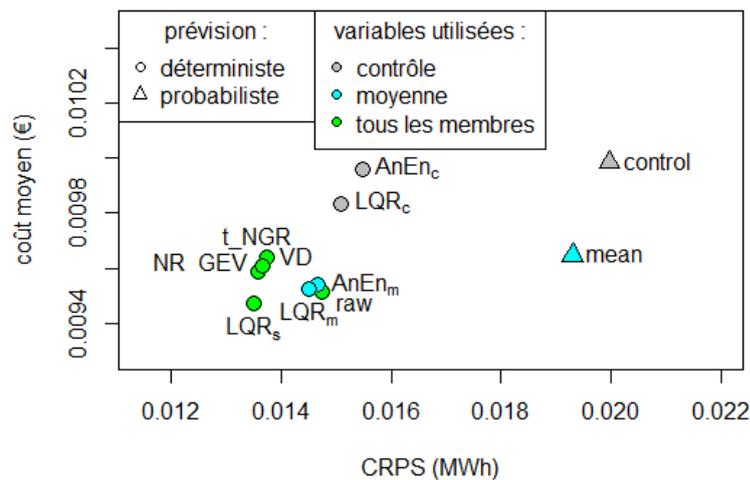


(c) Corrélation entre le coût moyen et le CRPS.

FIGURE 4.3.8 : Distribution uniforme.



(a) Illustration de la distribution centrée. (b) Corrélation entre le coût moyen et l'ignorance score.



(c) Corrélation entre le coût moyen et le CRPS.

FIGURE 4.3.9 : Distribution centrée.

La corrélation entre le CRPS et la valeur des modèles probabilistes est assez claire pour la distribution uniforme. Il se trouve que pour cette distribution particulière, l'équation 1.43 ($CRPS = 2 \times \int_{\tau=0}^{\tau=1} QS_{\tau} d\tau$) montre que le CRPS représente théoriquement exactement le double du coût moyen. Ce résultat généralise la capacité du Brier score à être un indicateur de la valeur pour les prévisions de variables binaires au CRPS pour les variables continues. En revanche, lorsque la distribution s'éloigne de cette distribution, la corrélation est moins évidente.

En règle générale, plus la distribution est "éloignée" de la distribution "centrée", plus la valeur des prévisions déterministes est faible. Pour la distribution centrée, les modèles probabilistes et déterministes ont un coût moyen très proche. Dans ce cas, l'approche probabiliste n'apporte pas de valeur, mais les modèles les plus performant sont ceux qui estiment correctement le niveau de probabilité 0.5. On voit alors apparaître une séparation

nette entre les modèles n'utilisant que le membre de contrôle, qui estiment relativement mal ce niveau de probabilité, et ceux qui utilisent la moyenne des membres. De son côté, l'*ignorance score* ne parvient pas à représenter correctement le coût moyen.

Une autre distribution particulière peut être simplement traitée théoriquement. Pour une distribution “décentrée”, pour laquelle $R = 0$ ou $R = 1$, le coût de toutes les prévisions probabilistes est nul, car la décision V optimale est triviale et vaut 0 ou ∞ . Le gain relatif des prévisions probabilistes par rapport aux prévisions déterministes est alors maximal. Pour ces cas extrêmes, où les distributions sont les plus éloignées de la distribution “centrée”, la valeur relative des prévisions probabilistes est maximale et le *CRPS* échoue alors à retranscrire la valeur des prévisions.

Le lien entre la valeur d'une prévision (pour des coûts de type “market-based”) et sa qualité évaluée par une métrique autre que le *quantile score* n'est finalement pas clairement établi. Lorsque la distribution des risques est uniforme, le *CRPS* quantifie correctement le coût d'un modèle, et donc sa valeur. Plus la distribution s'éloigne de cette dernière, moins le *CRPS* est un indicateur fiable de la valeur.

4.3.7 Liens avec le cadre d'évaluation des prévisions de variables binaires

Il est possible d'établir un parallèle entre les conclusions de ce travail théorique sur la valeur des prévisions probabilistes de variables continues et le cadre d'évaluation classique de la valeur des prévisions de variables binaires exposé dans la section 4.2. Ces liens sont mis en évidence par le tableau 4.3.2.

Type de variable	Ratio déterminant	Score adapté	Valeur potentielle
Binaire	C/P	<i>REV</i>	PEV
Continue	$\frac{P_+}{P_+ + P_-}$	<i>Quantile score</i>	PEVC
Type de variable	Valeur effective	Score dist. unif.	
Binaire	/	<i>Brier score</i>	
Continue	EVC	<i>CRPS</i>	

TABLEAU 4.3.2 : Outils d'étude de la valeur et liens entre qualité et valeur d'une prévision probabiliste : comparaison des prévisions de variables binaires et continues. “Score adapté” : score d'évaluation de la qualité modélisant correctement les coûts d'une prise de décision à ratio fixé. “Score dist. unif.” : score corrélé avec la valeur dans le cas où le ratio déterminant est uniformément distribué sur $[0;1]$.

4.4 Application à un cas pratique : Valorisation économique d'un parc de production solaire

Les méthodes de prévision considérées dans cette étude permettent de produire des prévisions à court terme performantes pour des horizons supérieurs à 12h. Or, l'essentiel des échanges d'électricité sur le marché australien se fait proche du temps réel (AEMO, 2020). Les modèles présentés ici ne semblent donc pas adaptés à l'optimisation de la vente de la production d'un parc sur le marché australien. C'est pourquoi le choix a été fait de simuler la vente d'électricité de ce parc solaire sur des marchés d'électricité d'autres pays :

le France, le Portugal et la Suisse. Ces pays ont été sélectionnés car les prix de vente sur ces marchés ont des caractéristiques assez différentes (cf. partie 4.4.3).

Cette section est une mise en application de la méthodologie présentée dans la partie 4.3. Les prévisions de production électriques présentées dans la partie 4.3.1 sont utilisées pour l'optimisation financière du parc de production considéré. Le fonctionnement réglementaire des marchés de l'électricité, ainsi que ses conséquences sur les prévisions et sur le contexte des coûts dans lequel évolue un acteur de marché sont détaillés dans la partie 4.4.1. Il met notamment en évidence la nécessité de disposer d'une méthode de prévision des prix pour cette optimisation financière. La partie 4.4.2 met en valeur l'importance du choix de l'objectif d'un utilisateur dans le processus d'aide à la prise de décision. Elle est suivie par la partie 4.4.3 qui présente le modèle de prévision des prix utilisé. Pour conclure, les résultats des simulations de ventes sur les différents marchés sont comparés et discutés dans la partie 4.4.4. Les outils de diagnostic présentés dans la partie 4.3 sont mis à profit pour interpréter les résultats finaux et discuter des différences entre les trois marchés.

4.4.1 Valorisation financière de la production d'électricité

Contexte et règles de marché

Le mouvement de libéralisation du secteur de l'électricité amorcé au début des années 2000 en Europe (Declercq et Vincent, 2000) a créé des marchés de l'électricité et a bouleversé les cadres réglementaires associés à la production d'électricité. Ces changements réglementaires ont entraîné des changements structurels avec l'apparition de nombreux acteurs privés. Parmi d'autres prérogatives, ces acteurs sont soumis à l'obligation d'équilibre, qui reflète une contrainte technique des réseaux électriques : à tout instant, la production électrique du réseau doit être égale à sa consommation pour maintenir la fréquence. Si cette contrainte est respectée par chacun des acteurs, l'équilibre global du réseau sera nécessairement respecté. Pour chaque acteur, au pas horaire², le gestionnaire du réseau calcule la somme des injections de cet acteur au réseau électrique (productions, achats, importations) et la somme des soutirages de cet acteur (consommations, consommations des clients éventuels, ventes, exportations). La différence entre ces deux grandeurs constitue le déséquilibre horaire de l'acteur. Ce déséquilibre fait l'objet d'une pénalité financière détaillée plus loin dans cette partie.

Cette libéralisation des marchés a par ailleurs été concomitante avec le développement des énergies renouvelables dans la plupart des pays européens. En France, un cadre réglementaire particulier a accompagné le début de ce développement. Cependant, depuis 2015, les producteurs d'énergies intermittentes sont dans l'obligation de valoriser par leurs propres moyens l'énergie produite. Ainsi, le déséquilibre d'un tel producteur qui valorise sa production sur les marchés "day-ahead" est égal à la différence entre sa production effective et sa vente sur la bourse de l'électricité. Le mécanisme de fonctionnement des bourses du marché de l'électricité est décrit ci-dessous.

²Sur quelques rares réseaux, le pas de temps du calcul peut être différent.

Fonctionnement des bourses du marché de l'électricité à J+1 et contraintes sur les prévisions

Dans le cadre d'un marché libéralisé et concurrentiel (tel que le marché européen depuis la fin du processus de libéralisation en 2009), un acteur du marché de l'électricité (courtier, fournisseur, producteur...) peut très rarement assurer par ses propres moyens son obligation d'équilibre. Souvent cet équilibre est trouvé *via* des échanges de volumes d'électricité entre acteurs. Un acteur uniquement producteur vend la totalité de sa production, quand un acteur uniquement fournisseur sera contraint d'acheter la totalité de la consommation de ses clients. Ainsi, des marchés d'échanges se sont naturellement développés. Les échanges d'électricité "day-ahead" (pour le jour J+1) peuvent s'opérer de 2 façons :

1. Soit via des contrats "over the counter" ("OTC"), aussi appelés "de gré à gré", qui sont des contrats directement négociés entre l'acheteur et le vendeur,
2. Soit via une bourse organisée gérant un "marché spot" (marché J+1).

La bourse du marché de l'électricité EPEX SPOT Day-Ahead gère depuis 2015 la bourse J+1 d'une grande partie de l'Europe de l'Ouest dont la France et la Suisse. Au Portugal, c'est le "MIBEL" ("Iberian electricity Market") qui gère la bourse d'échange J+1. Cependant, les marchés SPOT organisés par MIBEL et EPEX SPOT ont de nombreux points communs :

- Ces deux marchés sont dits de "Fixing", c'est-à-dire que les échanges n'ont pas lieu en continu, mais sont déterminés une fois par jour autour de 12h30 de la journée J pour toutes les heures de la journée J+1.
- Les acteurs souhaitant échanger des volumes d'électricité pour J+1 doivent faire parvenir leurs ordres d'achat et/ou de vente, au pas horaire, avant 12h00 au jour J. Un ordre d'achat (respectivement de vente) est constitué d'un volume (quantité d'électricité à échanger) et d'un prix associé, prix maximum (resp. prix minimum) auquel l'acteur est prêt à acheter (resp. à vendre) le volume d'électricité en question.
- Á 12h00 au jour J, une fois que tous les acteurs ont envoyé leurs ordres d'achats et/ou de ventes, la bourse calcule pour chacune des heures de la journée J+1 le prix d'échange d'électricité π^* appelé "prix spot" et le volume total d'électricité échangé selon la procédure décrite ci-après :
 1. Toutes les offres de vente sont triées par ordre de prix croissant. L'ensemble de ces offres triées peut être représentée sous la forme d'une courbe croissante dans le plan prix/quantité. Cette courbe est appelée "supply curve" (figure 4.4.1). De la même façon, les offres d'achat sont triées par ordre de prix décroissant pour former une "Demand curve",
 2. Le point d'intersection de la "Demand curve" et de la "supply curve" détermine à la fois le prix d'échange de l'électricité ("prix d'équilibre" ou "prix spot") et le volume échangé correspondant à ce prix d'équilibre (cf. figure 4.4.1),

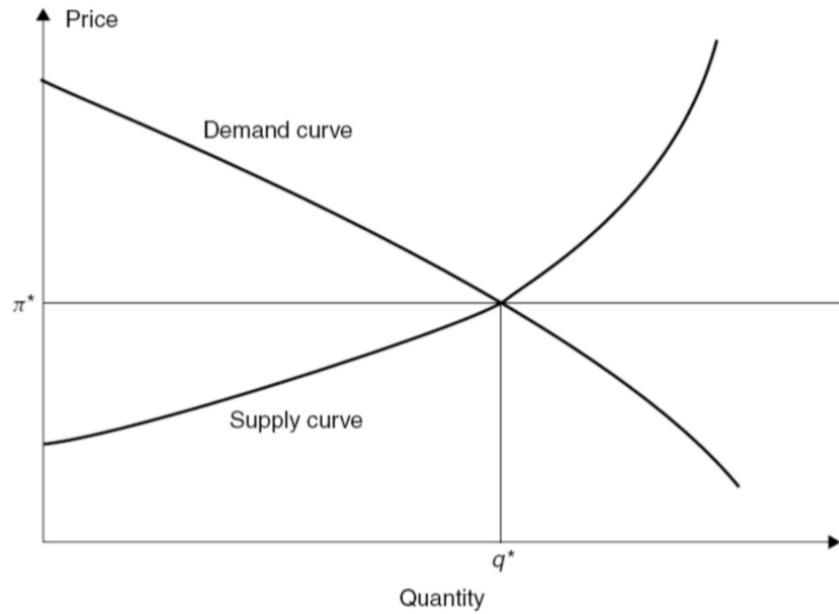


FIGURE 4.4.1 : Détermination du prix de marché et du volume échangé dans un marché à fixing. π^* : prix du marché ou prix spot ; q^* : volume échangé. Source : *Energy Markets*, Ernst (2016-2017).

3. L’algorithme de détermination des prix d’équilibre et des volumes d’échange est lancé indépendamment pour chacune des heures de la journée J+1. Les résultats correspondants déterminent les 24 prix d’équilibre π^* pour les 24 heures de la journée J+1,
4. Le processus temporel associé à ce fonctionnement est résumé sur la figure 4.4.2.

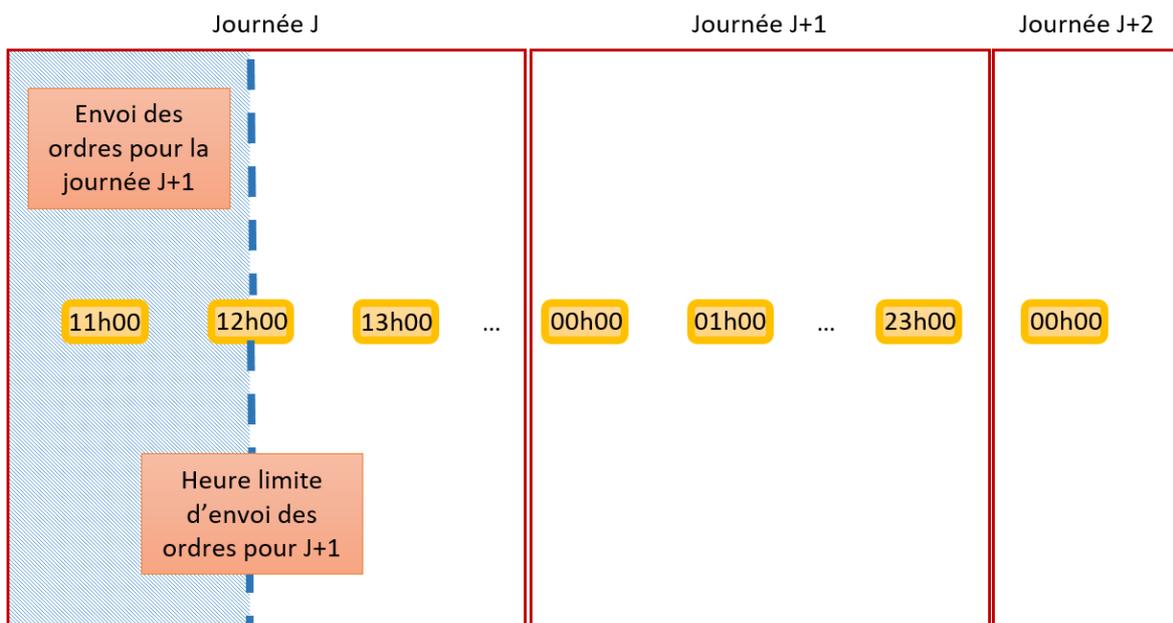


FIGURE 4.4.2 : Processus temporel du trading SPOT d’EPEX.

- Tous les acteurs ayant proposé des offres d’achats avec des prix supérieurs au prix π^* voient leurs ordres d’achats réalisés. De la même façon, tous les ordres de vente avec des prix inférieurs au prix π^* sont réalisés.

Dans ce mécanisme, chaque producteur qui souhaite valoriser sa production sur le marché doit déterminer son prix minimal de vente ; ce prix reflète son coût marginal de production. Comme tous les producteurs sont incités par ce mécanisme à produire des ordres d’achat dont les prix sont égaux à leur coût marginal de production, ce mécanisme trie naturellement les sources de production en fonction de ce coût marginal : c’est ce qu’on appelle le “merit order” (Cludius et al., 2014). La “supply curve” devient alors un classement des sources de production en fonction de leur coût marginal respectif, et seules les sources de production ayant un coût marginal inférieur à π^* sont nécessaires à l’équilibre offre-demande du système électrique. Pour la valorisation d’un parc de production d’énergie intermittente sans stockage, le coût marginal est nul. Ainsi, les sources intermittentes seront toujours classées premières de ce “merit order”. L’intérêt du producteur d’énergie intermittente est d’indiquer des prix de vente nuls pour être certain de vendre toute la production. Autrement, toute l’énergie non valorisée sur la bourse sera perdue (dans le cas simple de cette étude où le producteur ne dispose pas de moyen de stockage).

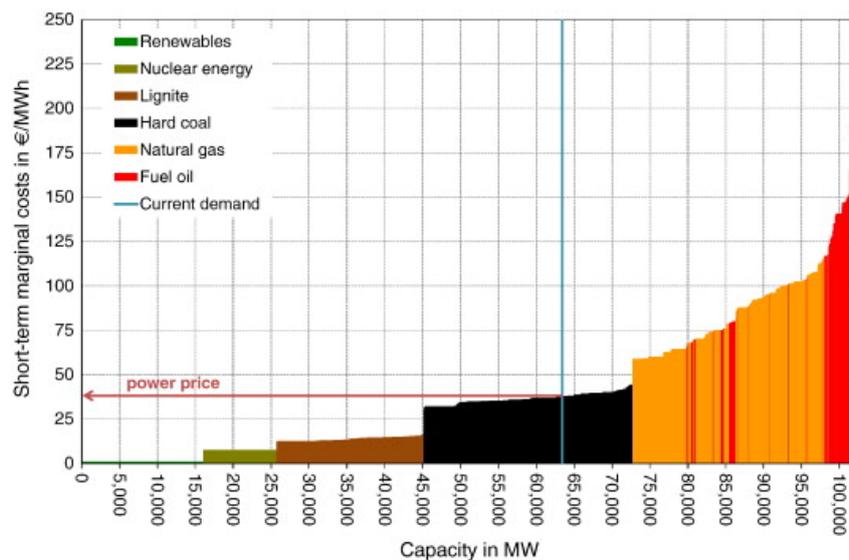


FIGURE 4.4.3 : “Merit order” typique de la “supply curve” d’un marché à fixing, source : *The merit order effect of wind and photovoltaic electricity generation in Germany 2008–2016 : Estimation and distributional implications*. Cludius et al. (2014).

Pour résumer, un producteur d’énergie solaire souhaitant valoriser sa production sur une bourse de l’électricité day-ahead peut tirer des mécanismes des marchés d’électricité présentés ci-dessus les 4 enseignements suivants :

1. Il doit proposer des offres de vente à 0 €,
2. Pour déterminer ses ordres de vente, il lui reste à déterminer à l’aide d’une prévision de sa production les volumes de ses offres de vente,
3. Les prévisions qui vont aider à déterminer les volumes des offres de vente pour toutes les heures de la journée J+1 doivent nécessairement être disponibles avant 12h00,

journée J. Autrement dit, une prévision couvrant toute la durée de la journée J+1 doit être disponible à 12h00 du jour J. Cette contrainte est illustrée par la figure 4.4.4.

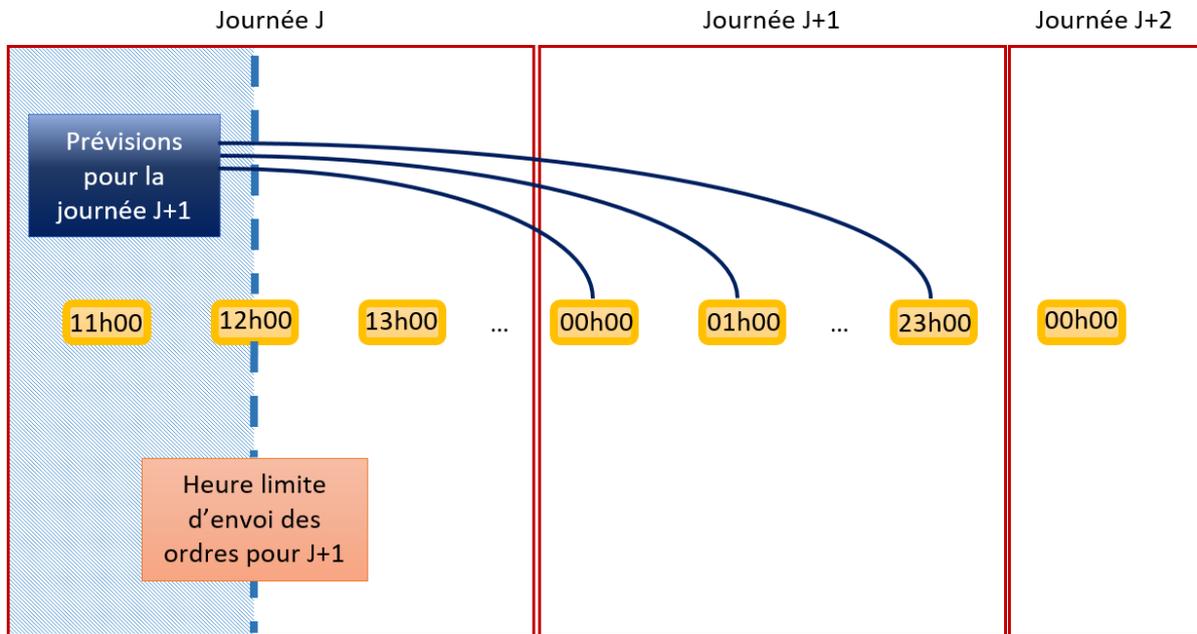


FIGURE 4.4.4 : Contraintes sur les prévisions utilisables pour générer des ordres de ventes sur les marchés J+1 d'EPEX SPOT ou de MIBEL.

4. Dans ce mécanisme, les revenus I tirés par le producteur de la vente d'un volume d'électricité E^c pour une heure donnée avec un prix spot π^* peuvent être calculés selon l'équation 4.20 :

$$R = E^c \times \pi^*. \tag{4.20}$$

Pénalités de déséquilibre

Pour inciter les acteurs privés à respecter la contrainte d'équilibre, le gestionnaire du réseau applique des pénalités financières à chaque acteur ne la respectant pas, à un pas de temps pré-déterminé (1h dans les cas de la France). De cette façon, les risques d'écarts entre la vente et la production effective sont supportés financièrement par le producteur. C'est ce mécanisme de rémunération qui s'applique également dans la plupart des pays européens. Cette pénalité financière T^* dépend en particulier du "sens" du déséquilibre (c'est-à-dire du signe de la différence entre la production et la consommation). Elle est déterminée par l'équation 4.21 :

$$T^* = \begin{cases} \pi^{*,+} \times D & \text{si } D > 0, \\ \pi^{*,-} \times |D| & \text{si } D < 0, \end{cases} \tag{4.21}$$

où $\pi^{*,+}$ et $\pi^{*,-}$ sont respectivement appelés pénalité de déséquilibre positif et pénalité de déséquilibre négatif et D est la différence entre la production effective E^* et la vente E^c :

$$D = E^* - E^c. \tag{4.22}$$

Les pénalités applicables au producteur peuvent donc être représentées graphiquement :

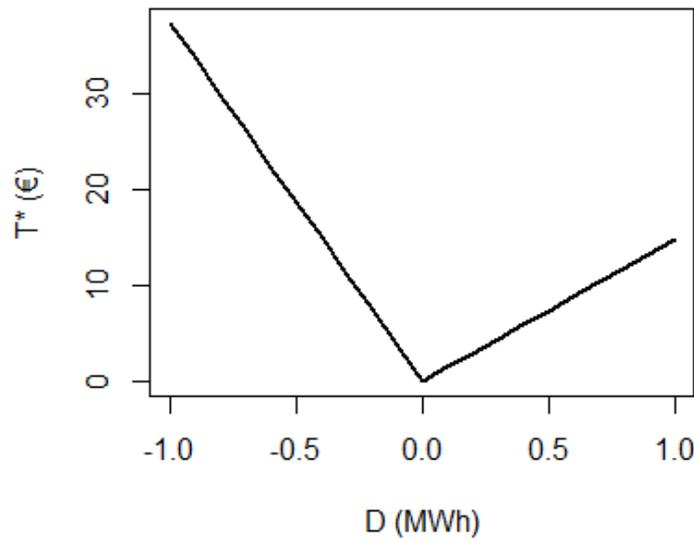


FIGURE 4.4.5 : Exemple de pénalités pour un producteur dans le cas où $\pi^{*, -} = 37.42$ €/MWh et $\pi^{*, +} = 14.78$ €/MWh (cas du marché suisse le 17/10/2015 14h00 heure locale).

Finalement, au pas horaire, le bénéfice réalisé par un producteur est déterminé par la différence entre ses revenus horaires (équation 4.20) et ses pénalités horaires (équation 4.21). 3 prix déterminent donc entièrement la rémunération de la production d'un parc d'énergie intermittente : le prix spot π^* , la pénalité de déséquilibre positif $\pi^{*, +}$ et la pénalité de déséquilibre négatif $\pi^{*, -}$. Pour une heure donnée, le bénéfice final est déterminé par l'équation 4.23 :

$$G = \begin{cases} E^* \times \pi^* - \pi^{*, +} \times D & \text{si } D > 0, \\ E^* \times \pi^* - \pi^{*, -} \times D & \text{si } D < 0. \end{cases} \quad (4.23)$$

Ainsi, les gains G d'un producteur peuvent être représentés en fonction de la déviation D :

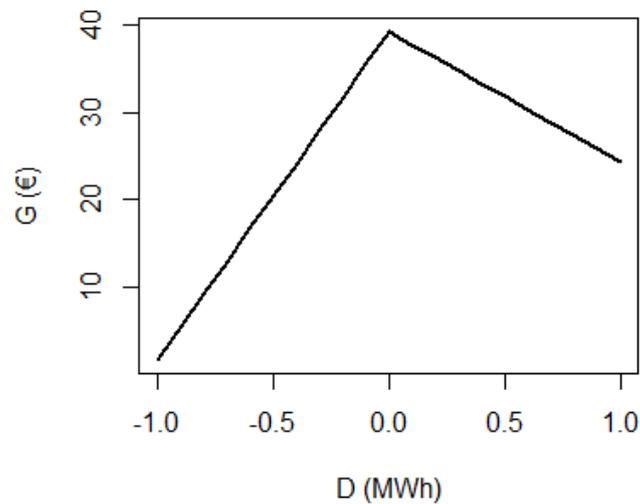


FIGURE 4.4.6 : Exemple de gains pour un producteur en fonction de la déviation D dans le cas d'une production effective E^* de 1 MWh avec : $\pi^*=39.18$ €/MWh ; $\pi^{*, -}=37.42$ €/MWh ; $\pi^{*, +}=14.78$ €/MWh (cas du marché suisse le 17/10/2015 à 14h00 heure locale).

4.4.2 Différentes stratégies d'optimisation

Une prévision est un outil d'aide à la prise de décision. Celle-ci doit être prise dans le but de réaliser un objectif défini en amont par l'utilisateur lui-même (par exemple : maximiser le gain, minimiser les probabilités d'obtenir une forte pénalité, maximiser l'espérance d'atteindre au moins un certain seuil de gain ...). La valeur d'une prévision doit alors être évaluée par une mesure de la réalisation, ou non, de cet objectif. Les prévisions probabilistes ont la propriété de pouvoir aider à maximiser l'espérance de la réalisation de cet objectif, *via* une optimisation. Au contraire, une prévision déterministe ne permet pas d'optimisation (le seul choix possible pour l'utilisateur est de suivre la prévision-point). L'évaluation de la valeur d'une prévision se fait ensuite par une mesure de la réalisation de l'objectif déterminé sur une période de test. Ce processus peut être résumé par le schéma 4.4.7.

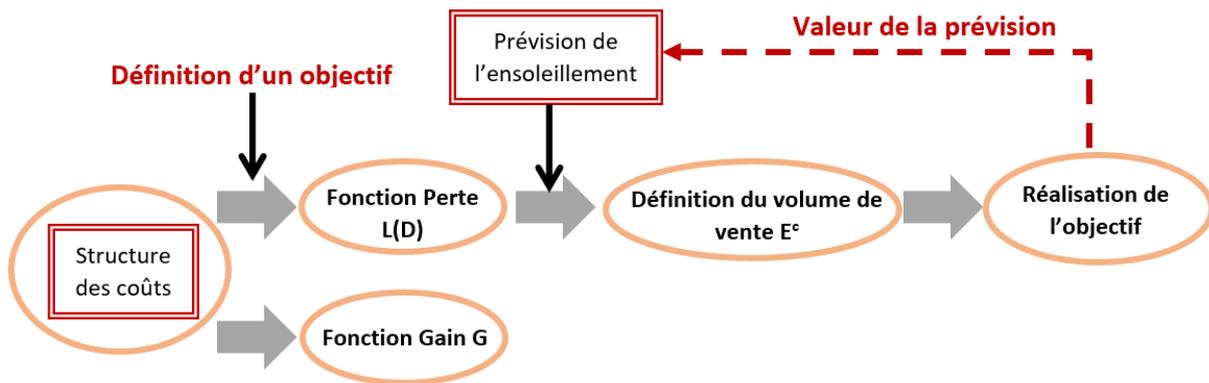


FIGURE 4.4.7 : Schéma du processus d'optimisation d'un moyen de production solaire en fonction de la structure des coûts et à l'aide d'une prévision de l'ensoleillement.

La structure des coûts permet de définir la fonction gain G . De plus, le choix d'un objectif par l'utilisateur d'une stratégie d'optimisation est modélisé par une fonction perte (Pinson et al., 2007), qui représente les coûts associés à une erreur de prévision. La prévision probabiliste de la production électrique et la fonction perte permettent la définition d'une fonction $\hat{\mathcal{P}}$, qui représente le coût attendu en fonction du volume vendu E^c , et qu'on cherche à minimiser (cf. section 4.3). Enfin, la valeur de la prévision probabiliste est mesurée par l'atteinte ou non de l'objectif.

Dans cette étude, nous avons souhaité étudier plusieurs fonctions pertes possibles dans le cas particulier de l'optimisation financière d'une centrale de production d'énergie photovoltaïque qui vend l'électricité produite sur le marché day-ahead. Dans chaque cas, la valeur de la prévision est associée à une stratégie d'optimisation définie par l'utilisateur. Plusieurs stratégies différentes ont pu être déterminées et étudiées (Shaomao, 2017). Ici, nous choisissons deux types de fonctions pertes correspondant à deux stratégies différentes : la maximisation du gain dans le cas d'un parc de production sans stockage (" O_1 ") et la maximisation du gain dans le cas d'un parc de production avec système d'appoint (" O_2 ").

Maximisation du bénéfice du parc sans stockage

Une première stratégie possible est de tenter de maximiser son profit. Cet objectif est appelé dans la suite de l'étude " O_1 ". Pour maximiser son profit sur une longue période, il faut maximiser l'espérance du profit pour chaque pas horaire, ce qui est équivalent à minimiser l'espérance des pénalités. C'est cette dernière formulation que nous allons retenir.

Dans la suite, la valeur d'une prévision pour cette optimisation est évaluée par la somme des gains accumulée par l'utilisateur. Pour cet objectif, la fonction perte $L(D)$ est exactement égale aux pénalités de déséquilibre T^* , comme montré sur la figure 4.4.8.

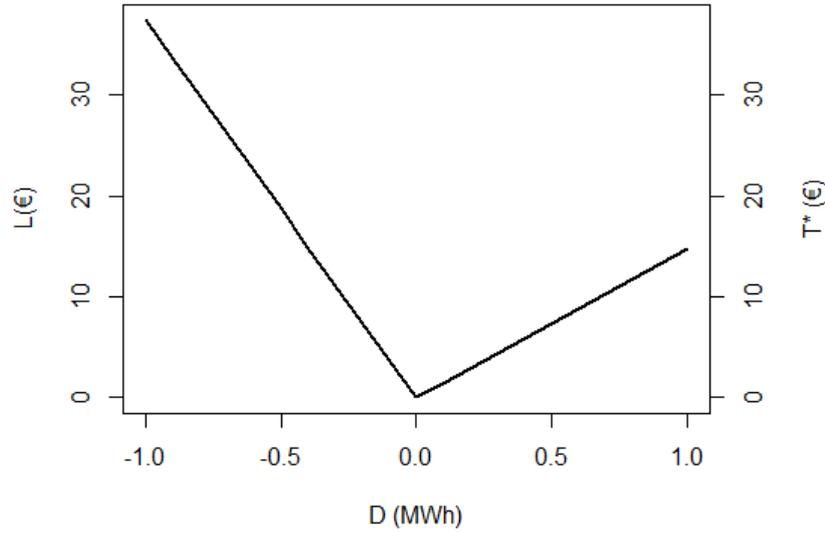


FIGURE 4.4.8 : Fonction perte $L(D)$ pour l’objectif O_1 dans le cas du marché suisse le 17/10/2015 à 14h00 heure locale où $\pi^{*,-}=37.42$ €/MWh et $\pi^{*,+}=14.78$ €/MWh. On a l’égalité $L(D) = T^*(D)$.

On constate que la fonction perte est de type “market-based” telle que définie dans la partie 4.3 avec par identification $\pi^{*,+} = P_+$ et $\pi^{*,-} = P_-$. Donc, pour cet objectif O_1 , la recherche faite dans la partie 4.3 peut être utilisée pour résoudre l’optimisation. De cette façon, pour une prévision probabiliste sous la forme d’une CDF $\hat{\mathcal{F}}$ de la production électrique, on sait que $\arg \min \hat{\mathcal{P}} = \hat{\mathcal{F}}^{-1}(R)$. Autrement dit, le volume de vente optimal est égal à $\hat{\mathcal{F}}^{-1}(R)$.

Une première approche vers le stockage

Si l’objectif O_1 est le plus couramment étudié dans la littérature (Matevosyan et Soder, 2006), il n’est pas le seul possible. Une fonction perte d’une forme différente peut servir à personnaliser la modélisation des coûts d’un utilisateur particulier. Par exemple, le cas de l’utilisateur avec une aversion aux risques qui souhaite minimiser la probabilité de voir apparaître une pénalité importante a déjà été étudié par Shaomao (2017).

Dans cette partie, nous allons choisir un objectif différent, afin de mesurer la valeur des prévisions probabilistes pour des fonctions pertes non “market-based”. Nous considérons l’exemple d’un utilisateur qui a la possibilité de stocker ou de soutirer une certaine quantité d’électricité δE_S à chaque pas horaire. Nous reprenons la modélisation des coûts d’un stockage proposée par Ramahatana (2021) où le coût d’injection dans le stockage est considéré nul. Nous avons arbitrairement considéré un coût de soutirage (prélèvement d’énergie dans le stockage) de 20 €/MWh. La fonction perte est donc une fonction affine par morceaux constituée de 4 segments : une pente de valeur $-\pi^{*,-}$ €/MWh pour $D = -\infty$ à $D = -0.1$, une pente de valeur -20 €/MWh pour D variant de -0.1 à 0 MWh, un segment de pente nulle entre $D = 0$ et $D = 0.1$ MWh et une pente de valeur $\pi^{*,+}$ €/MWh pour $D > 0.1$ MWh. Cet objectif prend dans la suite le nom “ O_2 ”. Une telle fonction de perte est représentée sur la figure 4.4.9.

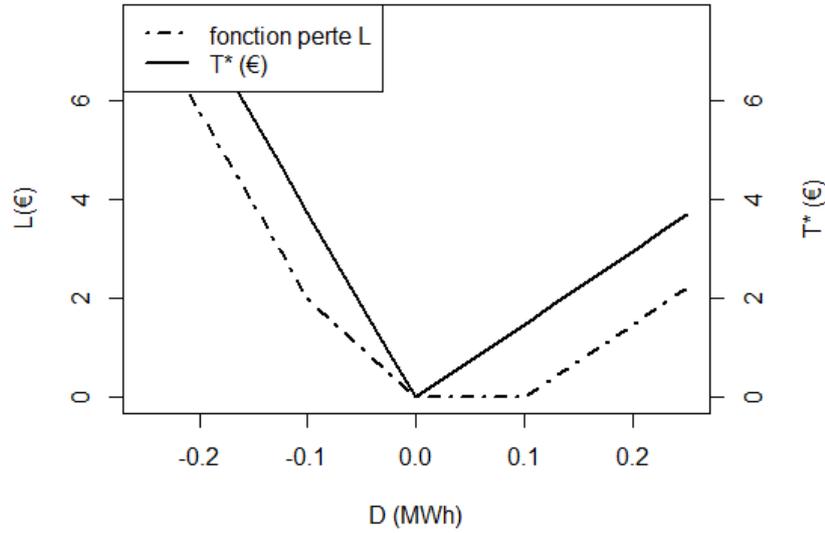


FIGURE 4.4.9 : Fonction perte L pour l’objectif O_2 dans le cas du marché suisse où $P_- = 37.42 \text{ €/MWh}$ et $P_+ = 14.78 \text{ €/MWh}$.

Dans le cas où la pénalité de déséquilibre négatif est inférieure à 20 €/MWh , l’utilisateur choisit de payer cette pénalité, et la fonction coût a alors une pente unique pour $D < 0$.

On peut noter le caractère fortement dissymétrique de cette fonction perte. Cette modélisation des coûts ne prétend pas modéliser l’optimisation d’un parc de production solaire avec stockage intégré. En effet, il faudrait prendre en compte à chaque pas de temps l’état de charge du stockage, son temps de réponse (Amaral Burghi et al., 2020), l’importance relative du prix spot de l’heure en cours par rapport aux prix des autres heures ou encore la dépendance temporelle des prévisions probabilistes³(Van der Meer et al., 2020; Golestaneh et al., 2016a). À la place, nous proposons ici une introduction vers les fonctions pertes de la forme de celles que rencontrerait un acteur exploitant un parc de production avec un système d’appoint (comme par exemple un stockage), dans le but d’étudier la valeur des prévisions probabilistes pour des fonctions pertes différentes des fonctions “market-based”. Par souci de simplification, dans la suite de l’étude, ce système d’appoint sera appelé “stockage”. Pour cet objectif O_2 , la résolution analytique de l’optimisation proposée dans la partie 4.3.3 ne s’applique plus car la fonction perte n’est plus de la forme “market-based”. Ainsi, la fonction $\hat{\mathcal{P}}$ est construite numériquement selon la procédure expliquée sur la figure 4.3.3.

Un exemple de fonction $\hat{\mathcal{P}}$ pour l’objectif O_2 , noté $\hat{\mathcal{P}}_2$, est présenté sur la figure 4.4.10. Elle est comparée avec $\hat{\mathcal{P}}$ de l’objectif O_1 ($\hat{\mathcal{P}}_1$). La fonction de prévision $\hat{\mathcal{F}}$ est aussi tracée. On constate que $\arg \min \hat{\mathcal{P}}_1 = \hat{\mathcal{F}}^{-1}(R)$. Ce n’est pas le cas pour $\hat{\mathcal{P}}_2$.

³Les dépendances temporelles peuvent être modélisées par des prévisions multivariées. Ce type de prévision n’est pas étudié ici.

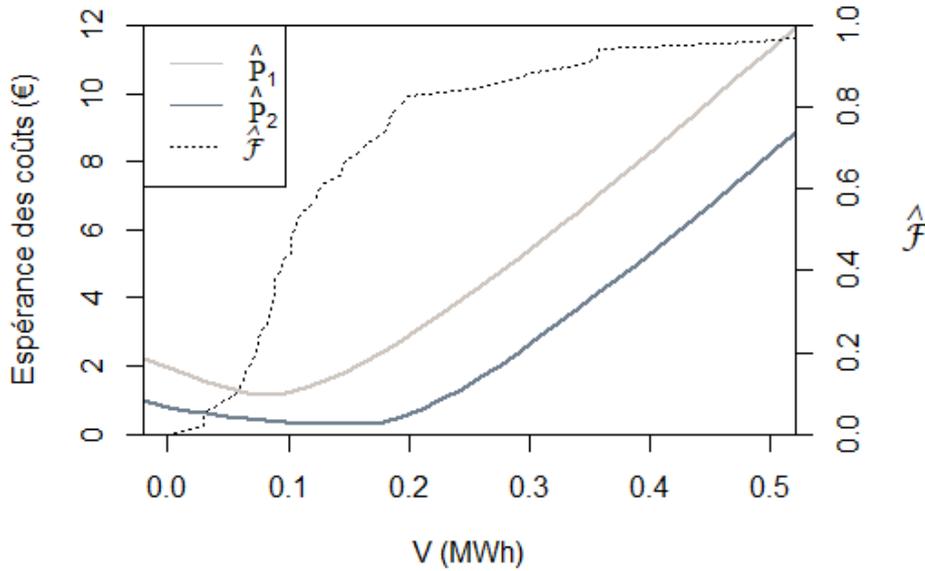


FIGURE 4.4.10 : Comparaison des fonctions \hat{P}_1 et \hat{P}_2 des objectifs O_1 et O_2 . Cas du marché suisse le 17/11/2017 à 07h00 heure locale, où $\pi^{*, -} = 27.79$ €/MWh et $\pi^{*, +} = 9.71$ €/MWh (prévision “VD”). Les volumes de ventes optimaux sont 0.08 MWh pour O_1 et 0.15 MWh pour O_2 .

4.4.3 Prévision déterministe des prix de déséquilibre

Les prix π^+ et π^- sont calculés en fonction de règles établies qui dépendent du pays et du gestionnaire de réseau. Ils sont aussi déterminés après le temps réel, et l'utilisateur ne dispose pas de ces prix au moment de l'optimisation du parc. Pour simuler un cas réel d'utilisation, il est donc nécessaire de disposer d'une méthode de prévision des prix, en plus de la prévision de production.

Sélection des variables de prévision

Par analogie avec la partie 4.3, nous posons $\gamma = \pi^{*, +} + \pi^{*, -}$. Comme démontré dans la partie 4.4.2, les optimisations pour les objectifs O_1 et O_2 dépendent plus directement du ratio $R = \frac{\pi^{*, +}}{\pi^{*, +} + \pi^{*, -}}$ et de la somme $\gamma = \pi^{*, +} + \pi^{*, -}$ que des prix $\pi^{*, +}$ et $\pi^{*, -}$. Le prix spot π^* n'intervient jamais dans le travail d'optimisation et il n'est donc pas nécessaire d'en faire une prévision. C'est pourquoi les variables cibles des prévisions sont le ratio R et la somme $\pi^{*, +} + \pi^{*, -}$.

Dans cette partie, nous présentons la mise en place d'une méthode de prévision du couple (R, γ) . Ce couple détermine ensuite complètement les prix $\pi^{*, +}$, $\pi^{*, -}$ avec les transformations :

$$\begin{cases} \pi^{*, +} = R \times \gamma, \\ \pi^{*, -} = (1 - R) \times \gamma. \end{cases} \quad (4.24)$$

Cette prévision des couples (R, γ) est effectuée indépendamment pour les 3 réseaux électriques considérés qui sont les réseaux français (de gestionnaire RTE), portugais (Redes Energéticas Nacionais), et suisse (Swissgrid). Par souci de simplification, une seule méthode de prévision est mise en place pour chaque réseau, et cette méthode est optimisée pour la prévision du ratio R (identifié comme la variable la plus importante). Cette même méthode de prévision est ensuite appliquée pour la prévision de γ .

Par ailleurs, dans toute cette partie, ces méthodes de prévision des prix supposent que l'influence de l'activité de l'utilisateur sur les prix du marché est nulle : il est ce qu'on appelle un "price-taker".

La méthode ARIMA

Le couple (R, γ) est en général assez difficile à prévoir, en raison notamment de l'absence de variables exogènes associées. Ces prix sont calculés selon des règles internes au gestionnaire de réseau, et dépendent souvent du déséquilibre global du réseau et du coût supporté par le gestionnaire de réseau pour résorber le déséquilibre global (RTE, 2020 ; ERSE, 2020 ; Swissgrid, 2020). Ces variables sont elles-mêmes *a priori* très difficiles à prévoir. Afin de trouver une stratégie de prévision adéquate, la littérature se rapportant à ce problème a été étudiée. Weron (2014) ou Jakaša et al. (2011) mettent en évidence que les modèles de type ARIMA sont des techniques appropriées et couramment utilisées. Ce type de prévision s'appuie sur les autocorrélations temporelles des variables cibles. Des tests de vérification de ces autocorrélations ont été menés et jugés satisfaisants pour utiliser cette technique. Ils sont présentés dans l'Annexe 4.D.

Cependant, on peut observer de fortes variations périodiques dans les autocorrélations, en particulier, l'heure de la journée apparaît clairement être d'une importance capitale (cf. table 4.D.1). Les recommandations de Contreras et al. (2003) qui présentent un modèle ARIMA de prévision des prix de l'électricité pour les marchés californiens et espagnols présentant aussi de fortes dépendances à l'heure de la journée ont été suivies. De plus, pour se rapprocher au plus près des contraintes réelles d'un utilisateur industriel, les contraintes opérationnelles s'appliquant à un utilisateur réel expliquées dans la partie 4.4.1 ont été respectées pour la prévision des prix. D'après les règles de marché, il faut envoyer chaque jour avant midi les prévisions de vente pour le lendemain. Ainsi, les valeurs disponibles pour les prévisions des prix du jour $J+1$ sont les valeurs produites avant midi du jour J . Cette considération a entraîné le choix d'utiliser un modèle ARIMA différent pour chaque heure de la journée.

Résultats des prévisions de prix

Dans cette partie, les résultats des prévisions des prix pour les trois marchés considérés sont présentés. Ils tiennent compte de toutes les obligations opérationnelles, et pourraient donc être réellement obtenus par un utilisateur industriel souhaitant effectuer l'optimisation financière de son parc de production.

Pour chaque pays, la figure 4.4.11 montre la comparaison des ratios R réels et des prévisions de ratios issues des méthodes de prévision ARIMA pour chacun des trois pays, pour les journées du 01/01/2017 et 02/01/2017. Cette figure permet d'apprécier les fortes disparités entre les dynamique des ratios R des différents pays. La relative stabilité du ratio en Suisse facilite la prévision. Au contraire, les ratios sont très variables en France

et au Portugal, mais c'est la forte importance de la variable de saisonnalité (l'heure de la journée) qui permet à la prévision de garder une bonne qualité.

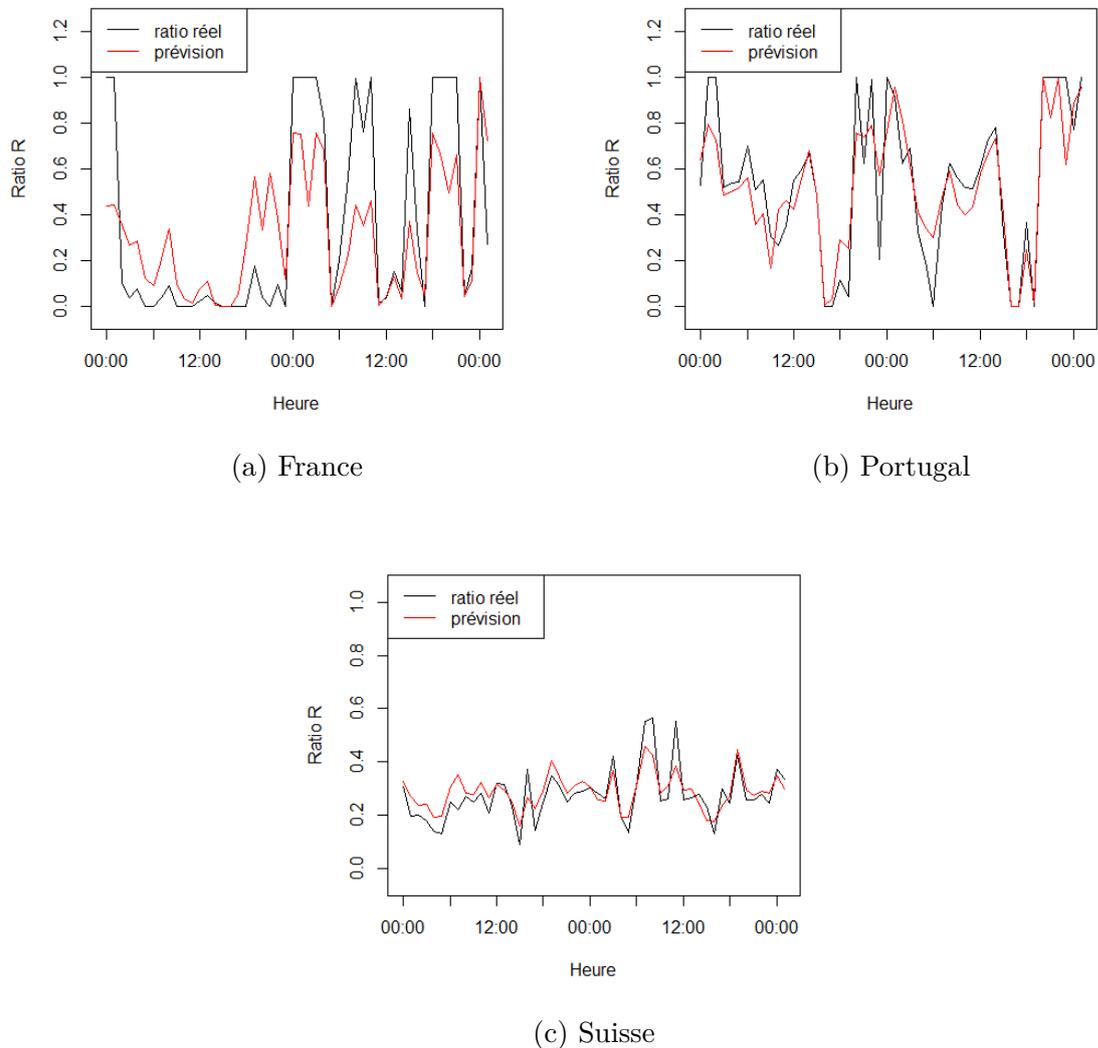


FIGURE 4.4.11 : Prévisions du ratio R pour les journées du 03 et du 04/01/2017.

Les RMSEs des prévisions des prix sont présentés dans le tableau 4.4.1. Les résultats des prévisions dépendent très fortement du pays considérés. En ce qui concerne R , qui a été identifié comme la variable la plus importante, la prévision du marché français est beaucoup moins performante que les prévisions des marchés suisse et portugais. De même, la prévision de γ sur le réseau suisse a un RMSE extrêmement élevé. Néanmoins, ces différences peuvent s'expliquer par la variance des variables cibles. La difficulté de prévision d'une variable ayant une forte variance est reconnue plus difficile, en raison du dilemme biais-variance (Woschnagg et Cipan, 2019). Les variances des différentes variables sont présentées sur le tableau 4.4.2. Les variables possédant un fort RMSE sur le tableau 4.4.1 correspondent effectivement aux variables ayant une forte variance sur la tableau 4.4.2.

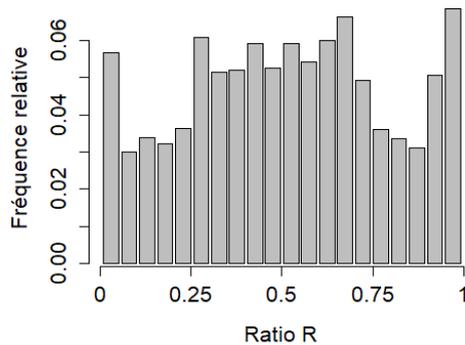
Pays	Ratio (R)	Somme (γ)
France	0.273	7.25
Portugal	0.132	9.98
Suisse	0.106	47.34

TABLEAU 4.4.1 : $RMSE$ des prévisions des variables de prix.

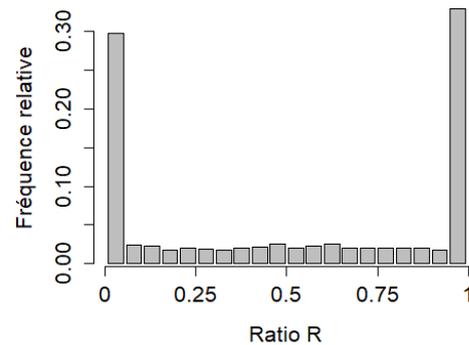
Pays	Ratio (R)	Somme (γ)
France	0.18	167.27
Portugal	0.06	269.12
Suisse	0.05	3309.86

TABLEAU 4.4.2 : Variance des variables de prix.

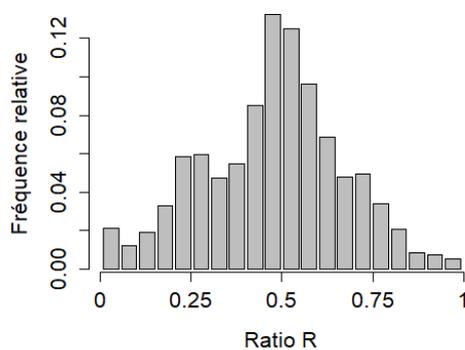
En plus des $RMSEs$, nous présentons sur la figure 4.4.12 une comparaison des distributions des ratios R réels et prévus pour les trois pays.



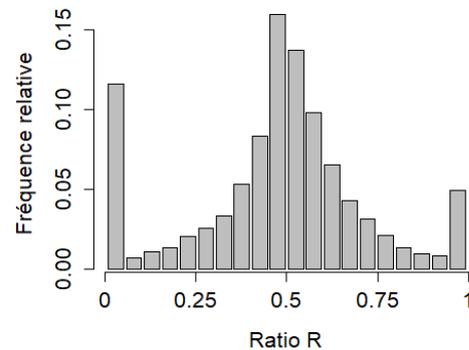
(a) Prévision du ratio R : France



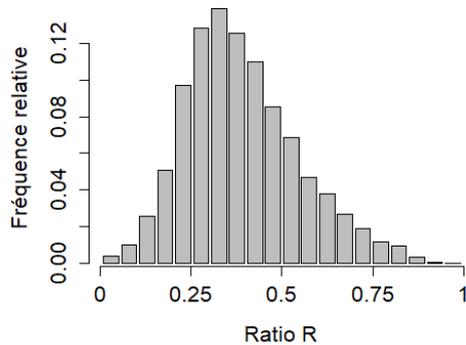
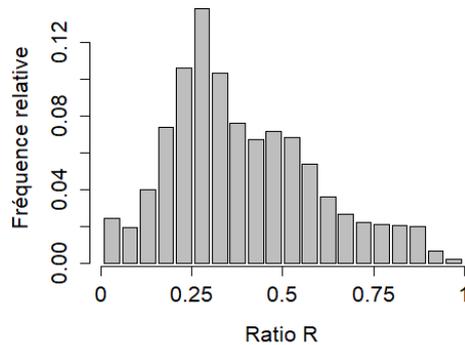
(b) Ratio R réel : France



(c) Prévision du ratio R : Portugal



(d) Ratio R réel : Portugal

(e) Prédiction du ratio R : Suisse(f) Ratio R réel : SuisseFIGURE 4.4.12 : Distribution du ratio R pour les 3 pays.

Les distributions de la prédiction sont assez différentes des distributions réelles, et en particulier plus centrées sur le ratio $R = \frac{1}{2}$ (notamment pour le réseau français). Cela peut s'expliquer par le fait que le modèle de prédiction ARIMA cherche à minimiser l'erreur quadratique et tend ainsi à donner des prévisions conservatrices et peu risquées, pour éviter les fortes erreurs de prédiction.

4.4.4 Résultats

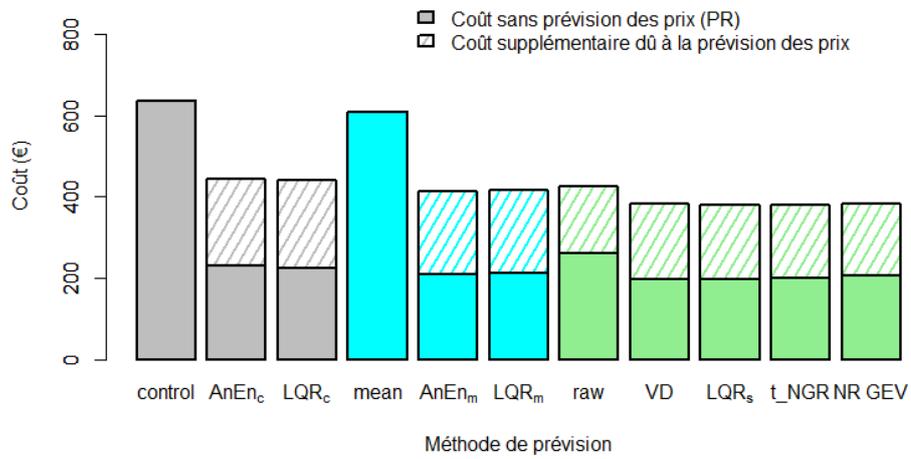
Cette étude a été faite en deux parties. Dans un premier temps, les optimisations ont été effectuées en prenant les couples (R, γ) réels comme structure de coût d'entrée des modèles d'optimisations (cf. schéma figure 4.4.7) : c'est le scénario "Prix Réels" ou "PR". Ce scénario permet de montrer les valeurs "brutes" des prévisions probabilistes seules, ce qui est le but principal recherché dans cette étude.

Cependant, cette application a également pour but de montrer des performances que pourraient obtenir un acteur utilisant réellement ces méthodes de prévisions sur un marché de vente d'électricité. Il est alors nécessaire de recourir à la prédiction des prix. C'est pourquoi les optimisations ont été faites une nouvelle fois en utilisant la prédiction des prix décrite dans la partie 4.4.3 (scénario "Prix prévus" ou "PP"). Dans ce scénario, les coûts en entrée des modèles d'optimisation ne sont plus les prix réels mais les prix fournis par les prévisions. Les résultats obtenus lors de ces 2 scénarios seront le plus souvent possible comparés entre eux pour distinguer les performances brutes des prévisions probabilistes et la perte de valeur supplémentaire due à l'incertitude associée à la prédiction des prix.

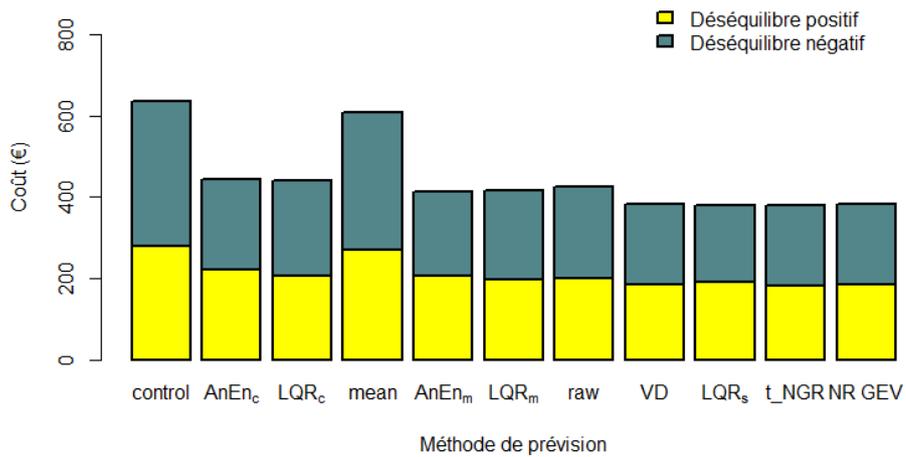
À la fin de la période de test, les bénéfices engendrés par chaque méthode de prédiction ont été additionnés. Ces bénéfices pour le marché français sont présentés sur le tableau 4.4.3. Le bénéfice maximal possible pendant la période de test compte-tenu de la production effective et des prix de marché spot π^* du réseau considéré est atteint par la prédiction parfaite (modèle parfait). Le coût associé à chaque modèle est la somme des pénalités négatives et positives. Ce coût est aussi égal à la différence entre les bénéfices du modèle parfait et ceux du modèle considéré.

Résultats pour les deux objectifs proposés

Deux graphiques différents sont proposés : le premier compare pour chaque modèle de prévision les coûts du scénario PP et les coûts du scénario PR. Les barres pleines indiquent les coûts des prévisions pour le scénario PR. La partie hachurée restante indique donc les pertes (en €) dues à l'imparfaite prévision des prix. Les résultats du scénario PP sont directement ceux que pourrait obtenir un industriel, car ils prennent en compte toutes les incertitudes identifiées dans le processus d'optimisation d'un parc de production solaire. Le deuxième graphique montre la décomposition des coûts finaux du scénario PP en pénalités de déséquilibre positif et pénalités de déséquilibre négatif. Dans un premier temps, les figures 4.4.13, 4.4.14 et 4.4.15 montrent les résultats obtenus pour l'objectif O_1 .

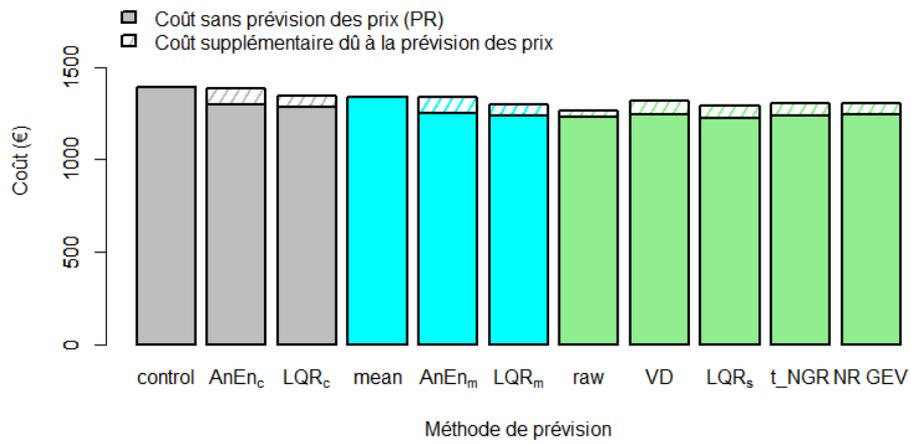


(a) Coûts des différents modèles avec et sans prévision de prix.

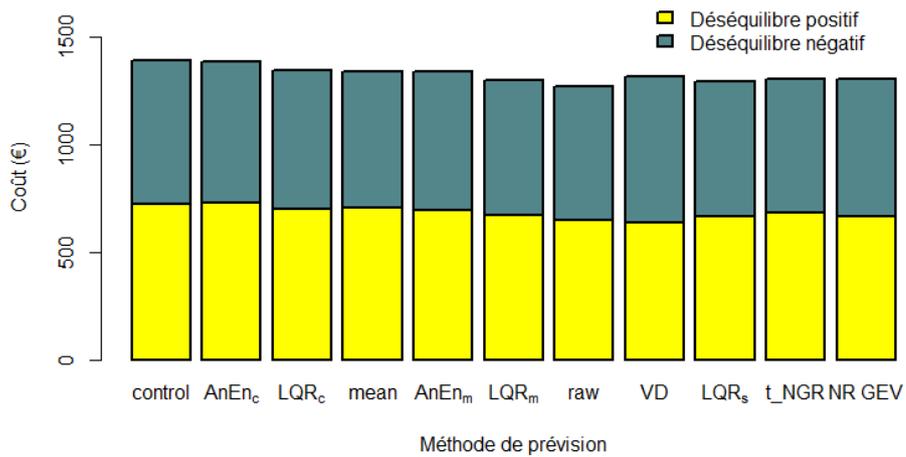


(b) Décomposition des coûts en fonction du sens du déséquilibre, scénario PP.

FIGURE 4.4.13 : Résultats pour l'objectif O_1 sur le marché français.

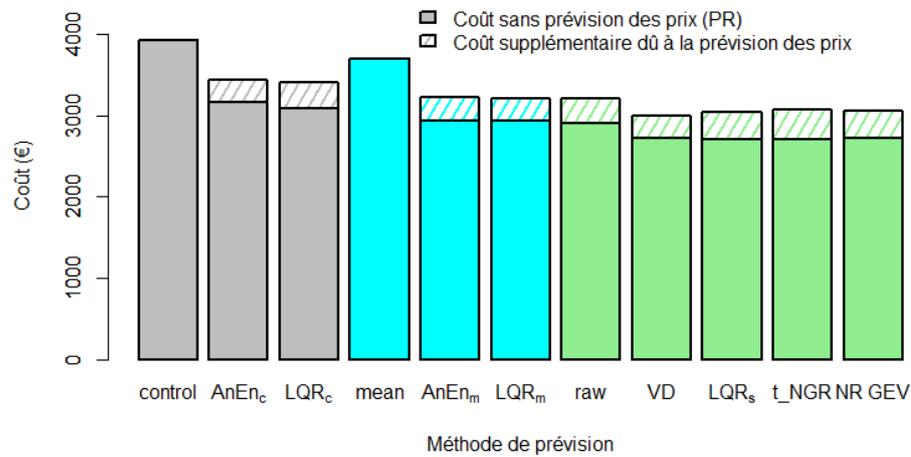


(a) Coûts des différents modèles avec et sans prévision de prix.

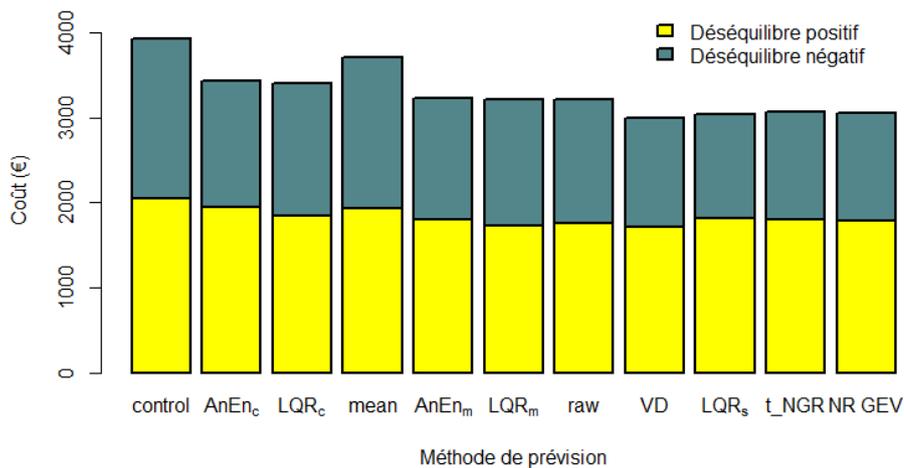


(b) Décomposition des coûts en fonction du sens du déséquilibre, scénario PP.

FIGURE 4.4.14 : Résultats pour l'objectif O_1 sur le marché portugais.



(a) Coûts des différents modèles avec et sans prévision de prix.



(b) Décomposition des coûts en fonction du sens du déséquilibre, scénario PP.

FIGURE 4.4.15 : Résultats pour l'objectif O_1 sur le marché suisse.

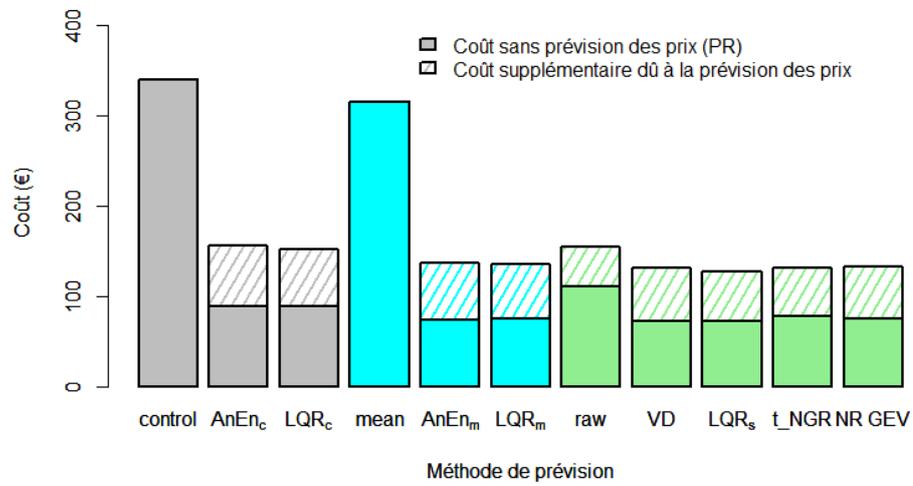
Dans les trois pays considérés, l'intérêt d'utiliser une prévision probabiliste est incontestable. En effet, les deux plus mauvais modèles sont toujours les prévisions déterministes. En revanche, le degré de supériorité des prévisions probabilistes dépend très fortement du pays.

La prise en compte des prévisions de prix comme données d'entrée des modèles d'optimisation augmente les coûts. Il est notable que les prévisions des prix ne modifient pas la valeur des prévisions déterministes (la valeur des prix n'influe pas sur l'optimisation de ces prévisions, cf. partie 4.3). Malgré cela, les prévisions déterministes restent les plus mauvaises prévisions pour optimiser la vente du parc photovoltaïque même pour le scénario PP. Aussi, même pour le scénario PP, il reste intéressant d'utiliser la meilleure prévision probabiliste (au sens de la qualité). Pour tous les pays, on peut constater que :

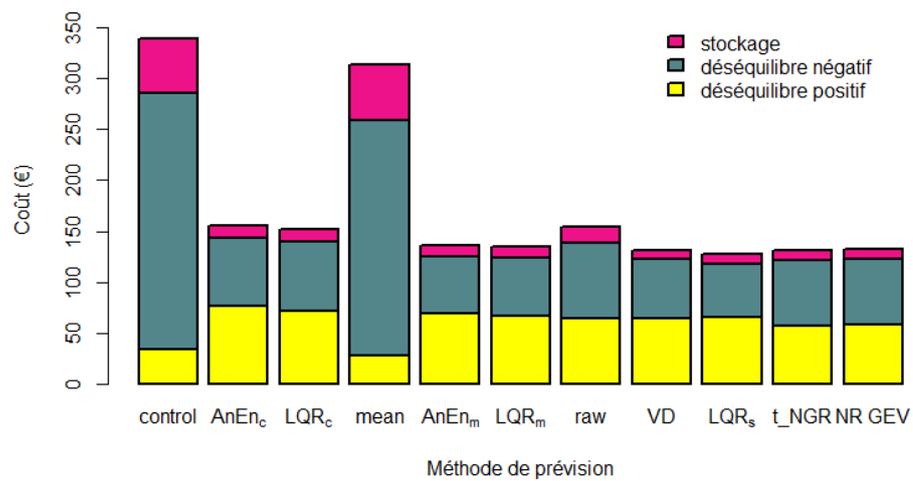
1. Les meilleurs modèles sont les modèles “Ensemble-based” utilisant tous les membres de l’ensemble brut, devant les modèles “Ensemble-based” n’utilisant que la moyenne des membres. Les modèles “Deterministic-based” sont toujours les plus mauvais,
2. La calibration des prévisions d’ensemble apporte un gain de valeur (excepté pour le cas du Portugal, qui est étudié en détail dans la section suivante à l’aide du graphique EVC).

La comparaison des résultats obtenus pour les différents réseaux permet de constater que le marché français est de loin le marché pour lequel les coûts supplémentaires dus à la prévision des prix sont les plus importants. Cela peut s’expliquer à la fois par le fait que la prévision des prix du ratio R est de moins bonne qualité en terme de RMSE (cf. tableau 4.4.1) et par le fait que les prix donnés par la prévision en France sont nettement plus centrés sur $\frac{1}{2}$ que les prix réels. (cf. figure 4.4.12). Les prises de décision issues des optimisations probabilistes se rapprochent alors des prévisions déterministes, et la valeur des prévisions probabilistes se rapproche alors d’autant de la valeur des prévisions déterministes.

Les résultats de l’optimisation O_2 sont présentés sur les figures 4.4.16, 4.4.17 et 4.4.18 sous la même forme que les résultats de O_1 .

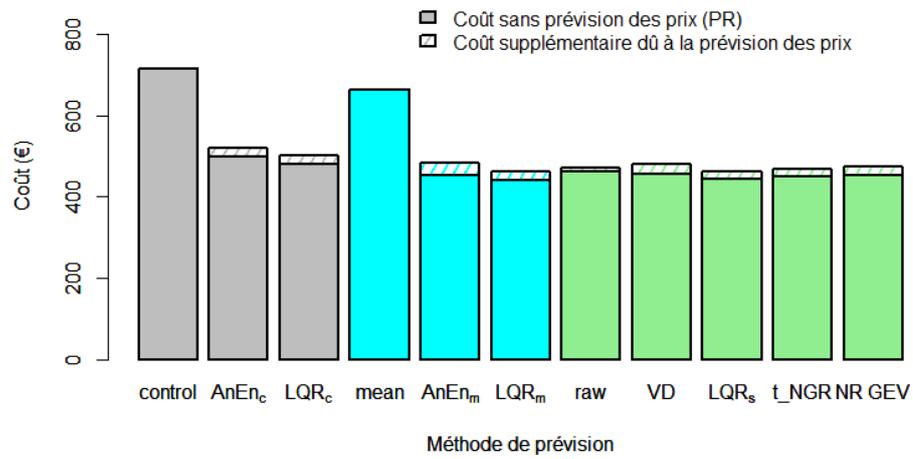


(a) Coûts des différents modèles avec et sans prévision de prix.

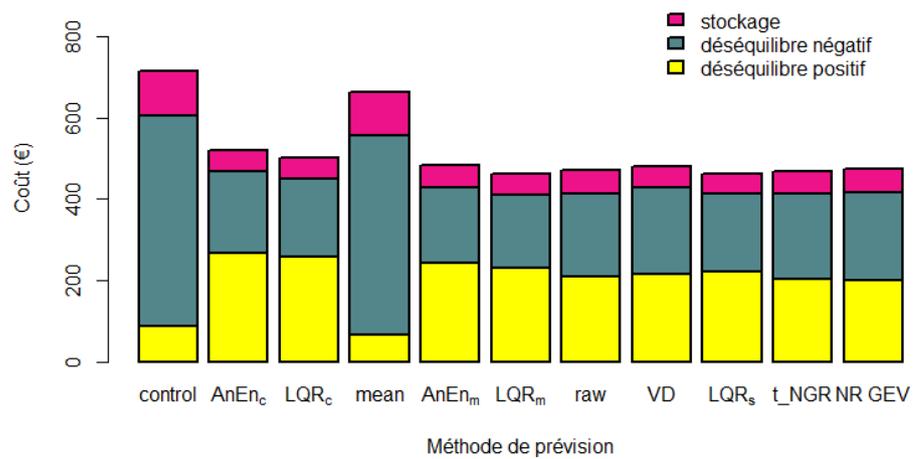


(b) Décomposition des coûts en fonction du sens du déséquilibre, scénario PP.

FIGURE 4.4.16 : Résultats pour l'objectif O_2 sur le marché français.

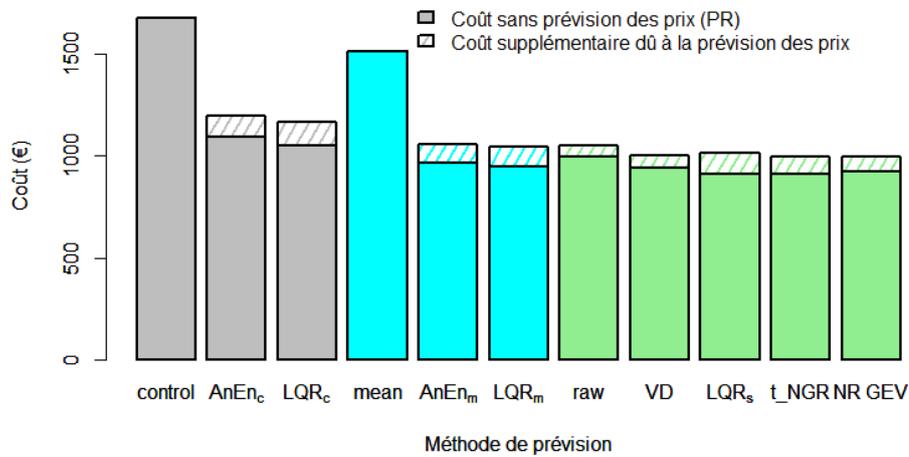


(a) Coûts des différents modèles avec et sans prévision de prix.

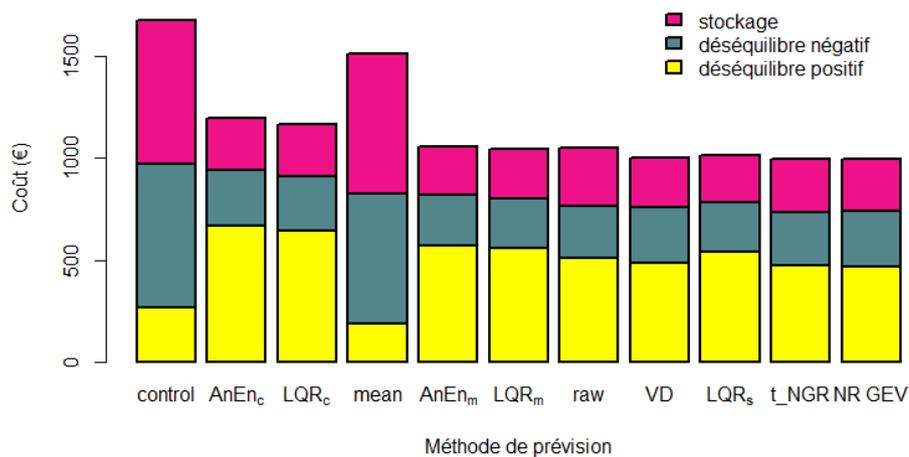


(b) Décomposition des coûts en fonction du sens du déséquilibre, scénario PP.

FIGURE 4.4.17 : Résultats pour l'objectif O_2 sur le marché portugais.



(a) Coûts des différents modèles avec et sans prévision de prix.



(b) Décomposition des coûts en fonction du sens du déséquilibre, scénario PP.

FIGURE 4.4.18 : Résultats pour l'objectif O_2 sur le marché suisse.

On constate que la différence des coûts entre PP et PR est moins importante que pour O_1 et cela est dû au fait que même si la prévision est imparfaite, les prix du stockage sont connus sans erreur à l'avance, ce qui diminue de fait l'importance des prix de déséquilibre dans l'optimisation.

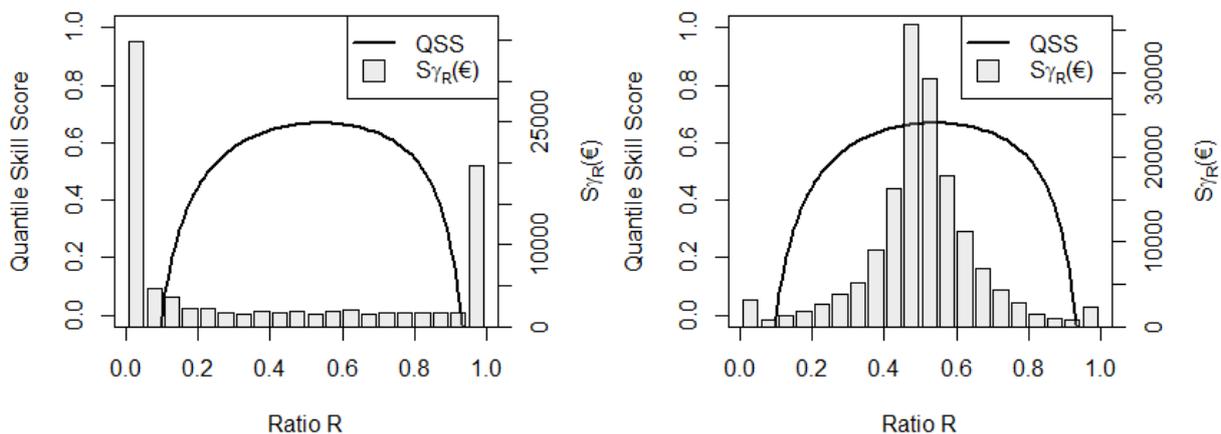
Par ailleurs, la différence de valeur entre les prévisions probabilistes et déterministes est nettement plus importante pour O_2 que pour O_1 . Cette augmentation de la valeur des prévisions probabilistes est due à l'option supplémentaire proposée par le stockage dont peuvent profiter les optimisations probabilistes. Les décisions optimales déterminées par celles-ci s'éloignent des médianes des fonctions de prévision PDF, et donc des prévisions déterministes. On peut constater que les optimisations probabilistes parviennent à s'adapter au changement de contexte en équilibrant les coûts de déséquilibre négatifs

et positifs. À l'inverse, pour prévisions déterministes, les coûts de déséquilibres négatifs surpassent très largement les coûts de déséquilibre positifs.

Explication des résultats par le graphique EVC

La section précédente a montré qu'il n'y a pas de réelles différences de valeur pour la distribution de ratios R la plus centrée autour de $\frac{1}{2}$ (le Portugal). À l'inverse, pour une distribution de ratios très extrêmes (cas de la France), les prévisions probabilistes ont un coût beaucoup plus faible (une valeur beaucoup plus élevée).

La valeur des modèles dans le scénario PR peut être expliquée à l'aide de la méthodologie présentée tout au long de la partie 4.3. Pour analyser les résultats du scénario PR, la figure 4.4.19 présente le graphique EVC appliquée à ce cas pratique. L'explication des résultats du scénario PP est plus complexe car la prévision des prix est un paramètre externe à l'optimisation qui a pourtant une influence forte sur le résultat final.



(a) Prévision de contrôle : marché français

(b) Prévision de contrôle : marché portugais

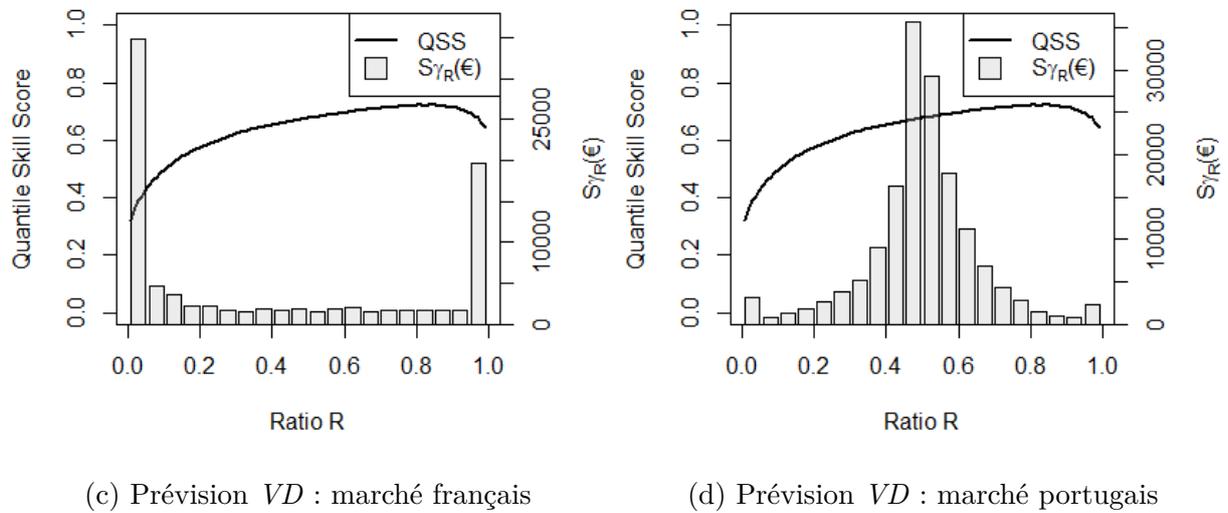
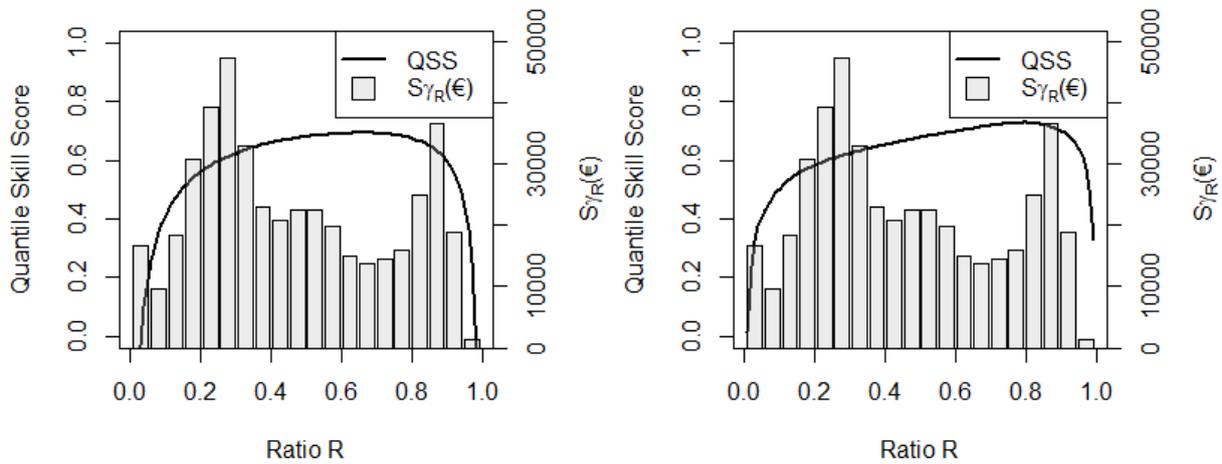


FIGURE 4.4.19 : Graphiques EVC des prévisions “control” et VD pour les marchés français et portugais.

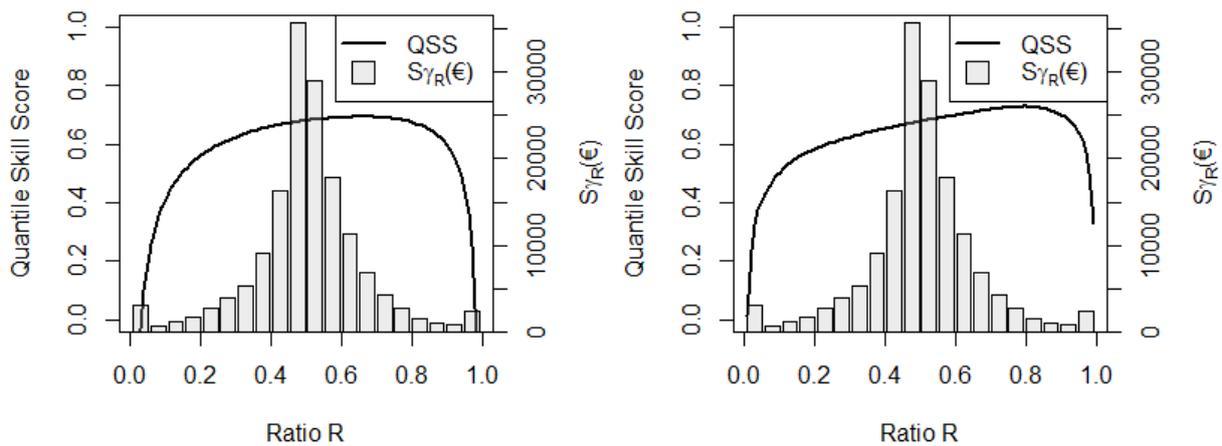
De tels graphiques expliquent très rapidement et simplement pourquoi la différence de valeur entre les prévisions déterministes et probabilistes (pour le scénario PR) est si forte en France et si faible au Portugal. Les prévisions déterministes sont des prévisions très inégales, au sens où leur QSS n’est fort que pour un intervalle faible de ratios R , là où au contraire, les prévisions probabilistes ont des valeurs de QSS relativement constantes. Il se trouve que les ratios les plus importants pour le réseau portugais coïncident avec les ratios pour lesquels les prévisions déterministes ont une valeur importante. Cela explique que la supériorité des prévisions probabilistes en terme de qualité ne se traduit pas par un gain fort en terme de valeur. En revanche, le phénomène inverse peut être observé dans le cas du réseau français. Les ratios les plus importants sont les ratios extrêmes, ce qui avantage fortement les prévisions probabilistes.

Dans la partie 4.4.4, il a été démontré que la calibration des prévisions probabilistes apportait une valeur nulle ou presque à la prévision pour une vente sur le marché portugais. On essaie dans cette partie d’expliquer ce phénomène grâce aux graphiques EVC combinant PEVC et répartition des risques S_{γ_R} . Pour cela, la figure 4.4.20 montre ces graphiques pour les prévisions brutes ECMWF et le modèle de calibration $NR\ GEV$.



(a) Prévision ECMWF brute : marché suisse

(b) Prévision *NR GEV* : marché suisse



(c) Prévision ECMWF brute : marché portugais

(d) Prévision *NR GEV* : marché portugais

FIGURE 4.4.20 : Graphiques EVC des prévisions ECMWF brutes et *NR GEV* pour les marchés suisse et portugais.

Comme on peut le voir, l'effet de la calibration sur les graphiques EVC est principalement d'améliorer la qualité de la prévision pour les niveaux de probabilité extrêmes. Alors que certains niveaux de probabilité extrêmes ont une importance forte sur le marché suisse, la prévision de ces niveaux est très peu importante dans le contexte de prix du marché portugais. Cela explique que la calibration a peu d'effet sur la valeur des prévisions appliquées au Portugal.

Décomposition des coûts

Le tableau 4.4.3 résume les performances des différents modèles sur le marché français en mettant l'accent sur la somme des volumes de déséquilibre constatés durant la période de

test et le coût moyen du déséquilibre, défini par le coût total associé à une méthode de prévision divisé par le volume total de déséquilibre.

Modèle	Vente (MWh)		Bén. (€)		Dés. (MWh)		C.M.D (€/Mwh)	
	PR	PP	PR	PP	PR	PP	PR	PP
climato	933	799	31 970	31 324	913	708	1.35	2.76
control	761	761	32 047	32 047	157	157	4.66	4.66
AnEn _c	768	750	32 377	32 164	533	312	1.21	2.64
LQR _c	736	749	32 382	32 167	412	265	1.34	2.73
mean	759	759	32 062	32 062	152	152	4.49	4.49
AnEn _m	734	742	32 397	32 194	483	289	1.21	2.63
LQR _m	724	743	32 396	32 191	387	250	1.36	2.84
raw	741	757	32 346	32 182	218	184	2.73	3.55
VD	749	761	32 409	32 225	371	262	1.25	2.61
LQR _s	710	739	32 409	32 227	335	235	1.40	2.88
t_NGR	771	765	32 406	32 228	315	227	1.44	2.83
NR GEV	776	764	32 401	32 225	353	244	1.47	2.80
parfait	774	774	32 606	32 606	0	0	/	/

TABLEAU 4.4.3 : Tableau récapitulatif des performances des différentes prévisions : exemple du marché français. *Vente* : Total de l'énergie vendue ; *Bén.* : Bénéfice total ; *Dés.* : Total de l'énergie de déséquilibre ayant fait l'objet de pénalités ; *C.M.D.* : Coût moyen de la pénalité de déséquilibre ; *PR* : Optimisations faites à partir des prix réels (sans prévision) ; *PP* : Optimisations faites à partir de la prévision des prix.

On peut noter que certaines caractéristiques semblent communes à tous les modèles probabilistes. La minimisation des coûts par ces prévisions ne passe pas par une diminution du volume d'énergie de déséquilibre. Au contraire, leurs volumes de déséquilibre sont plus élevés. En revanche, les prévisions probabilistes parviennent à diminuer le coût moyen du déséquilibre. C'est donc bien grâce à la possibilité d'intégrer les prix de déséquilibre dans leur processus d'optimisation que les prévisions probabilistes se démarquent. On voit d'ailleurs que la dégradation de la valeur des prévisions probabilistes dans le scénario PP se traduit par une augmentation de ce coût du déséquilibre moyen, alors que le volume de déséquilibre diminue. Les optimisations probabilistes sont alors plus proches des optimisations déterministes. Cela s'explique notamment par le fait que les prévisions du ratio pour le marché français sont plus conservatrices et centrées sur $R = \frac{1}{2}$ que les prix réels, ce qui rapproche les décisions issues des optimisations probabilistes des prévisions déterministes. Ces résultats suggèrent que l'importance de la qualité de la prévision des prix est cruciale pour ce type d'optimisation.

Importance de la prévision des prix

Dans cette partie, l'importance de la prévision des prix sur les coûts finaux subis par l'utilisateur des prévisions est discutée. Pour chaque modèle, la somme des pertes dues à l'imperfection de la prévision des prix a pu être identifiée dans la partie 4.4.4. Pour analyser plus précisément ce phénomène, les données de la période de test ont été regroupées en fonction du niveau de l'erreur de prévision du ratio R (scénario PP). Pour chacune des

catégories ainsi créées, les pénalités moyennes ont été calculées pour les deux scénarios. La figure 4.4.21 montre la population relative de chaque catégorie, et compare les pénalités moyennes en fonction de l'erreur de prévision du ratio R du scénario PP pour la prévision ECMWF brute.

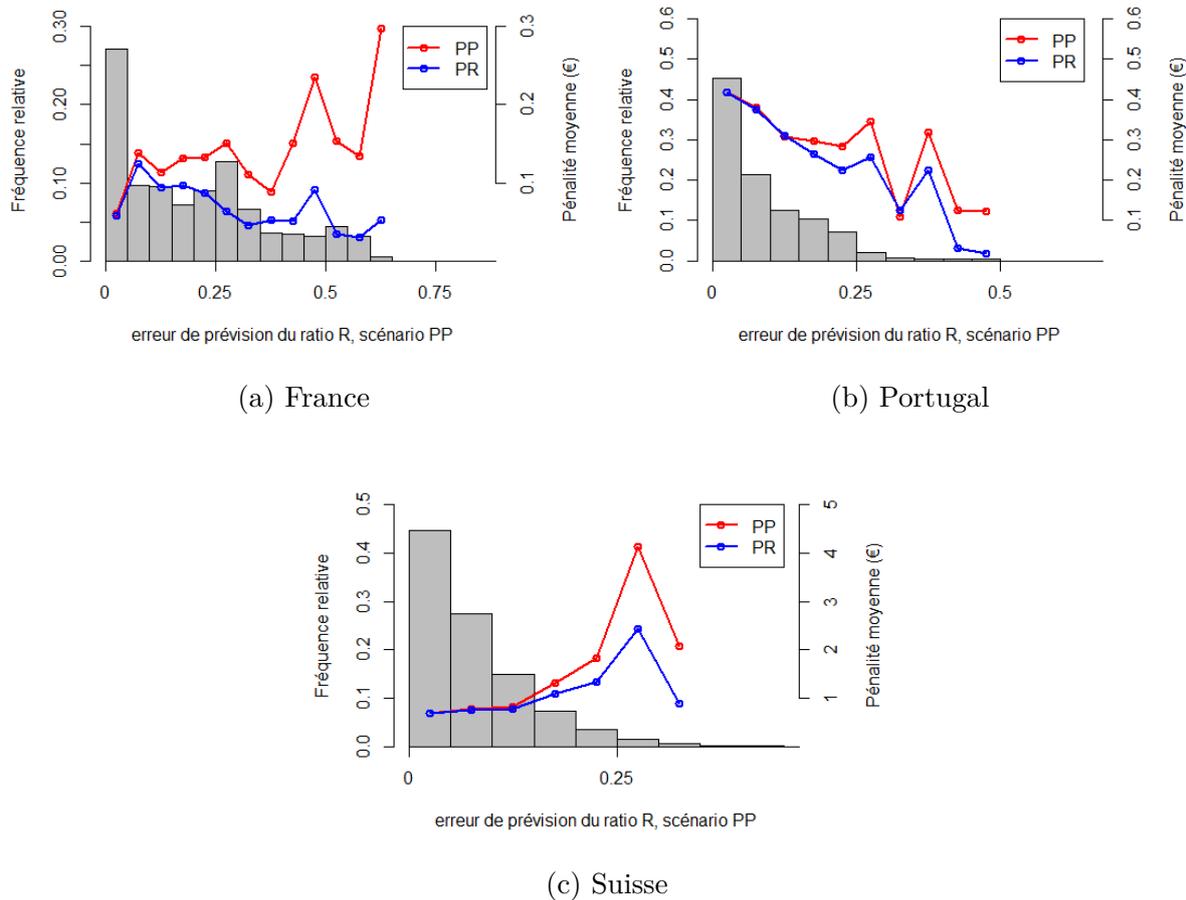


FIGURE 4.4.21 : Pénalité moyenne en fonction de l'erreur de prévision du ratio R pour le scénario PP.

La figure 4.4.21 montre clairement que quel que soit le pays, les pénalités moyennes sont plus importantes dans le cas où l'optimisation est faite à partir de la prévision des prix. En outre, plus l'erreur de prévision du ratio R augmente, plus la différence entre les pénalités moyennes des 2 scénarios augmente. Ce phénomène est particulièrement marqué lorsque l'erreur de prévision du ratio dépasse 0.25. Il existe cependant certaines exceptions qui peuvent s'expliquer par la très faible population des catégories concernées.

Influence de la prévision des prix sur le lien entre qualité et valeur

Pour le scénario PP, la corrélation entre le *quantile score* des prévisions et les pénalités a été calculée en fonction de l'erreur de prévision du ratio. Le résultat est montré sur la figure 4.4.22

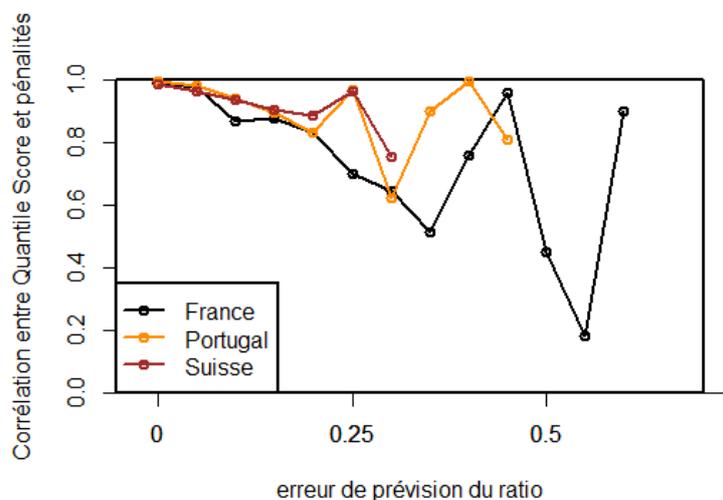


FIGURE 4.4.22 : Corrélacion entre le *quantile score* et la valeur des prévisions en fonction de l'erreur de prévision des prix (scénario PP).

La figure 4.4.22 montre que la corrélation est de 1 pour une erreur de prévision nulle. C'est une conséquence directe des résultats de la partie 4.3. La corrélation se dégrade ensuite lorsque l'erreur de prévision des prix augmente tout en restant faible. Lorsque celle-ci dépasse 0.2, il n'est plus possible de conclure quant à une corrélation entre le *quantile score* (mesure de qualité) et la valeur d'une prévision.

Gains effectifs de valeur des approches probabiliste et ensembliste et liens avec les gains de qualité

De manière analogue aux gains effectifs de qualité calculés dans le chapitre 3 (section 3.8.2), les gains effectifs attribuables à l'utilisation de l'approche probabiliste pour le scénario PR ont été calculés et les résultats obtenus sont résumés sur le tableau 4.4.4. Pour comparaison, le gain maximal de qualité de l'approche probabiliste pour le site d'Adélaïde, calculé à l'aide du MC-CRPS selon la procédure suivie dans le chapitre 3, est de 30.4 %.

Modèle	Marché français	Marché portugais	Marché suisse
AnEn _c	64.5	11.1	24.4
LQR _c	66.0	15.1	27.1
AnEn _m	66.2	10.9	24.6
LQR _m	66.3	13.4	25.7
raw Ensemble	57.1	7.8	21.5
VD	68.3	7.5	28.1
LQR _s	68.0	12.5	29.8
t_NGR	67.3	7.6	27.9
NR GEV	67.0	8.4	28.5

TABLEAU 4.4.4 : Gains effectifs (en %) de l'approche probabiliste par rapport à l'approche déterministe. Les valeurs en gras indique les meilleurs gains pour un marché donné.

On peut noter une grande diversité des gains de l’approche probabiliste par rapport à l’approche déterministe. Ces gains sont à mettre en relation avec les graphiques de répartition des risques des trois marchés concernés (cf. figures 4.4.20 et 4.4.19). En effet, d’après les conclusions de la partie 4.3.6, les gains spécifiquement attribuables à l’approche probabiliste sont plus importants lorsque la répartition des risques est éloignée de la distribution centrée.

Finalement, la figure 4.4.23 résume les gains maximums obtenus par l’approche probabiliste, comparativement à l’approche déterministe, lors des différents chapitres de ce travail. Il semble opportun de rappeler que les gains maximums de qualité ont été mesurés à l’aide du *CRPS*, et que les gains maximums théoriques (ou potentiels) ont été calculés en travaillant uniquement sur des distributions gaussiennes. Le gain effectif maximal pour un site est le gain effectif le plus important parmi tous les modèles implémentés dans ce travail. Ces gains effectifs sont du même ordre de grandeur que les gains théoriques maximums.

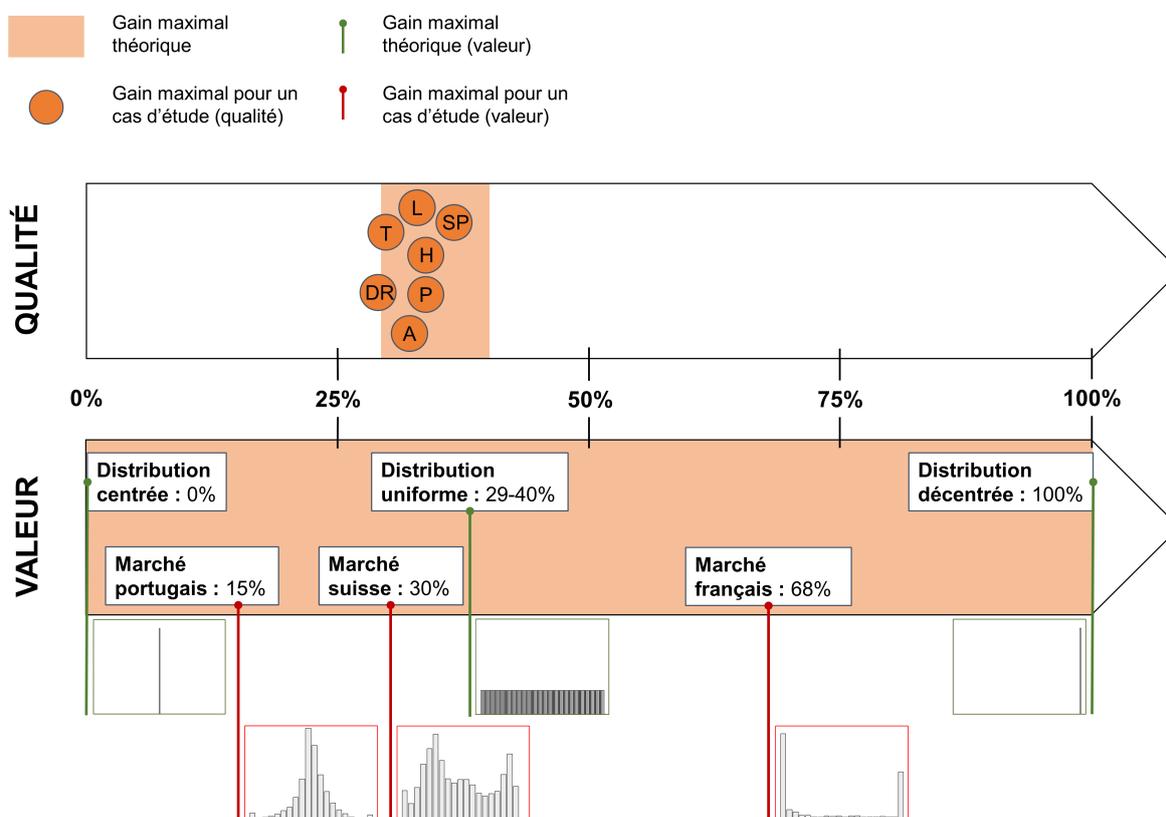


FIGURE 4.4.23 : Résumé des gains probabilistes maximums de valeur et de qualité de l’approche probabiliste. “DR” : Desert Rock, “T” : Tiruvallur, “L” : Langley, “H” : Hawaii, “P” : Palaiseau, “SP” : Saint-Pierre, “A” : Adélaïde. Le cas d’étude considéré pour les calculs de gains effectifs de valeur est celui du site d’Adélaïde, pour lequel le gain en *CRPS* est de 30.4%.

De la même façon, les gains de l’approche ensembliste, c’est-à-dire attribuables à l’utilisation des prévisions d’ensemble pour générer les prévisions probabilistes (cf. figure 3.8.2), ont été déterminés suivant la procédure décrite dans le chapitre 3. L’ensemble de ces gains est résumé dans le tableau 4.4.5 et peut être comparé aux gains de qualité

calculés dans le chapitre 3.

Site	Qualité	Marché	Valeur
	Gain maximum		Gain maximum
Adélaïde	11.5	Français	14.0
Hawaii	3.7	Portugais	5.7
Desert Rock	11.9	Suisse	11.7
Saint-Pierre	7.8		
Palaiseau	16.4		
Tiruvallur	4.3		
Langley	16.2		

TABLEAU 4.4.5 : Gains (en %) de qualité et de valeur de l’approche ensembliste. Les gains de valeur ont été calculés à partir des prévisions du site d’Adélaïde.

Les gains en valeur et en qualité sont proches, ce qui laisse à penser que l’amélioration des prévisions grâce à l’approche ensembliste est distribuée assez équitablement pour tous les niveaux de probabilité. Autrement dit, la distribution des risques ne semble pas avoir une influence majeure sur les gains apportés par l’approche ensembliste.

4.5 Conclusions

Dans ce chapitre, la notion de valeur d’une prévision a été discutée. Les valeurs de différentes prévisions déterministes et probabilistes ont été comparées pour le cas de l’optimisation financière d’un parc de production photovoltaïque sur les marchés de l’électricité. Dans un premier temps, une brève revue bibliographique ayant pour sujet les méthodes d’évaluation de la valeur des prévisions probabilistes de variables binaires est proposée. En ce qui concerne les prévisions de variables continues, une recherche théorique générale sur la valeur des prévisions dans le cadre des structures de coûts dissymétriques a été menée. Dans cette recherche théorique, il a été établi que la métrique de qualité la plus appropriée pour rendre compte de la valeur d’une prévision est le *quantile score*. Dans le cas de fonctions de perte de type “market-based”, ce score décrit correctement la valeur d’un modèle de prévision. À partir de cette constatation, un nouvel outil de mesure de la valeur a été proposé : le graphique EVC. Le lien avec d’autres scores tels que le *CRPS* ou l’*ignorance score* a été discuté. Un parallèle a été établi entre les valeurs des prévisions probabilistes de variables binaires (dilemme coût-perte) et de variables continues (cas des fonction pertes “market-based”) *via* le tableau 4.3.2.

Dans un deuxième temps, une application réelle d’optimisation d’un parc de production a été menée. Pour mieux interpréter les résultats finaux, et pour se rapprocher des résultats réels que pourraient obtenir un industriel utilisant ces modèles de prévisions, cette application a été faite en 2 étapes :

1. La première, sans faire de prévision des prix et en prenant en compte les prix réels (cas théorique inapplicable en pratique, mais qui permet de comprendre les performances des modèles de prévision),
2. La deuxième, en construisant une prévision des prix respectant les contraintes des marchés de l’électricité européens. Elle reflète alors les résultats que pourraient

obtenir en pratique un acteur industriel.

La comparaison des valeurs des prévisions probabilistes et déterministes montrent un net avantage en faveur des prévisions probabilistes. Les différences de valeur entre modèles probabilistes et déterministes dépendent très fortement de la structure des coûts qui s'applique au processus de prise de décision (matérialisée par le ratio R et la répartition des risques γ dans l'étude). Le graphique EVC a été utilisé pour montrer le lien entre différence de valeur et structure des coûts. Enfin, il a été établi que le lien entre valeur et qualité est moins fort, dès lors que la structure des coûts est elle-même incertaine, et qu'elle doit aussi faire l'objet d'une prévision nécessairement imparfaite.

L'importance de la qualité de la prévision des prix a très nettement été mise en valeur dans ce travail. De plus, il a été montré que les variables de prix sont difficilement prédictibles. Il semble qu'une prévision probabiliste des prix pourrait pallier à cette inhérente imprédictibilité. Une étude combinant l'utilisation de prévisions probabilistes à la fois pour la production électrique et pour les variables associées au prix semble être la suite logique de ce travail.

Annexes

4.A La méthode “Variance Deficit”

Le modèle “Variance Deficit” (ou “VD”) est une méthode dite d’“ajustement membre-à-membre”. Cette famille de modèles utilise une approche spécifique au sens où elle modifie les valeurs des membres d’un ensemble brut, de telle sorte que l’ensemble résultant soit correctement calibré. Les modèles d’ajustement membre-à-membre transforment donc un ensemble non-calibré (E_1, \dots, E_M) en un ensemble calibré (E'_1, \dots, E'_M) :

$$(E_1, \dots, E_M) \rightarrow (E'_1, \dots, E'_M). \quad (4.25)$$

Une fois l’ensemble calibré, une CDF peut être construite en utilisant les transformations montrées par Lauret et al. (2019) (cf. section 3.2) si nécessaire.

Une technique très commune d’ajustement membre-à-membre est appelée “Variance deficit” (VD) (Alessandrini et al., 2013; Sperati et al., 2016). Elle a été imaginée spécifiquement pour corriger la variance de l’ensemble brut, qui est connue pour être très souvent sous-estimée par les prévisions d’ensemble brutes (cf. partie 3.B). Une méthode de prétraitement visant à supprimer le biais a donc été appliquée avant la méthode VD proprement dite. La procédure ci-dessous explique l’implémentation de ce modèle.

Tout d’abord, l’ensemble brut $(E) : (E_1, \dots, E_M)$ de moyenne μ_E est modifié via une transformation logit en un autre ensemble (E^*, \dots, E_M^*) :

$$E^* = \log\left(\frac{k}{1-k}\right), \quad (4.26)$$

où :

$$k_i = \frac{E_i - \max(E)}{\max(E) - \min(E)}. \quad (4.27)$$

Ensuite, l’idée de l’algorithme de Variance deficit est de définir un coefficient V_{def} qui représente le manque de variance de l’ensemble. Il est calculé à l’aide de la période d’apprentissage par :

$$V_{def} = \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N (\mu_{E_n^*} - O_n^*)^2}}{\frac{1}{N} \sum_{n=1}^N \sigma_n^*}, \quad (4.28)$$

où σ^* est l’écart-type de l’ensemble. En pratique, le coefficient V_{def} quantifiant le manque de variance de l’ensemble brut est souvent supérieur à 1. Un nouvel ensemble intermédiaire $(E_1^{*'}, \dots, E_i^{*'}, \dots, E_M^{*'})$ est calculé par :

$$E_i^{*'} = \mu_{E^*} + V_{def} \times (E_i^* - \mu_{E^*}), \quad (4.29)$$

puis l’ensemble de prévision calibré (E') donné par la méthode VD est finalement déduit de l’ensemble $(E^{*'})$ grâce à la transformation inverse du logit.

4.B Modélisation de la production photovoltaïque

Dans le cadre de cette étude, les prévisions d'ensoleillement sont utilisées pour construire des prévisions de la production d'électricité d'un parc de production photovoltaïque. Leur valeur est évaluée dans un contexte d'aide à la prise de décision pour la valorisation financière de ce parc. La prévision utilisée pour ce travail de modélisation est une prévision d'ensemble d'énergie de rayonnement fournie par ECMWF au pas de 3 heures. Cette prévision est tout d'abord transformée en prévision d'éclairement énergétique moyen incident pour chaque pas de 3h, ce qui constitue la prévision d'ensemble ($E_i^{GHI}, i = 1...51$). Elle est constituée d'un membre de contrôle et de 50 membres perturbés (pour lesquels les conditions initiales et certains paramètres du modèle sont légèrement modifiés). Ensuite, trois transformations successives sont appliquées à cette prévision brute pour obtenir une prévision probabiliste de la production électrique :

1. Etape 1 : Un algorithme de suréchantillonnage permettant d'obtenir une prévision d'ensemble (Eo_i^{GHI}) de 51 membres au pas horaire. Cette étape est détaillée dans la partie 4.B.2.
2. Etape 2 : Un algorithme mathématique de modélisation de la production électrique des panneaux photovoltaïques considérés, permettant la conversion de l'éclairement énergétique en production électrique. Cet algorithme est déterministe et transforme un membre Eo_i^{GHI} de l'ensemble (Eo_i^{GHI}) en une valeur de production E_i^p , pour former l'ensemble (E_i^p), qui est donc toujours constitué de 51 membres. La section 4.B.3 explique l'algorithme utilisé ici.
3. Etape 3 : Un modèle de post-traitement qui permet la calibration de l'ensemble (E_i^p). Ce modèle est une application qui transforme une prévision ensembliste non calibrée en une prévision probabiliste sous la forme d'une distribution de probabilité cumulée (CDF) ou d'un ensemble calibré. Les différents modèles de post-traitement utilisés sont présentés dans la partie 4.3.1.

L'ensemble de ce processus est résumé par la figure 4.B.1.

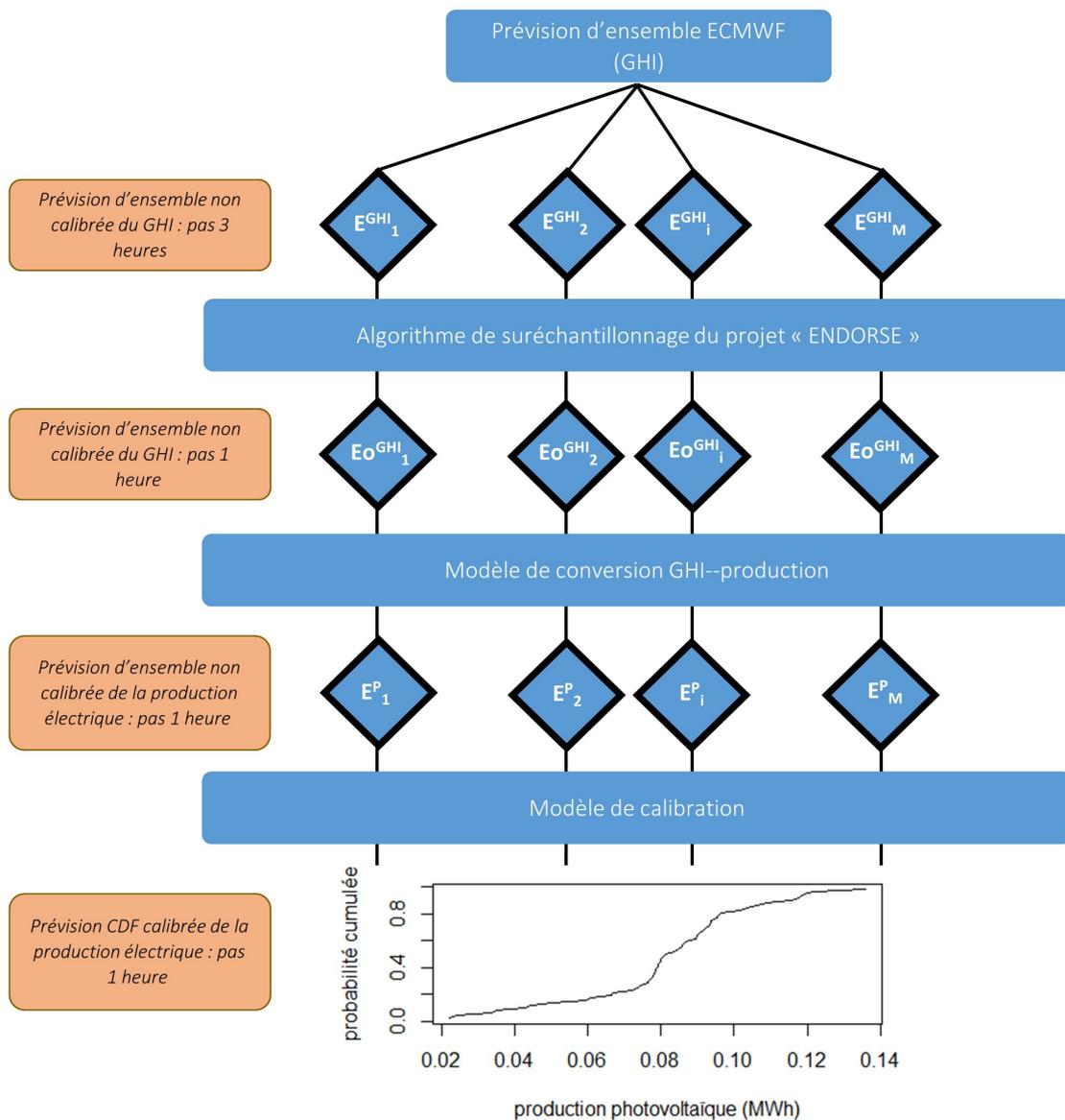


FIGURE 4.B.1 : Processus de modélisation de la production.

4.B.1 Données et présentation du parc de production

Les données de production photovoltaïque utilisées dans cette étude ont été obtenues par simulation à partir de mesures de GHI sur le site de l'aéroport d'Adélaïde, en Australie (-34.95°S, 138.52°O). Il est constitué de modules photovoltaïques, tous considérés horizontaux, dans le but de simplifier la modélisation. Ce sont tous des modules LG NEON2, de surface unitaire $1.64m^2$. Les spécifications techniques de chaque panneau sont résumés dans le tableau. 4.B.1.

Spécification	Valeur
Type	Monocristallin
Surface (m^2)	1.64
Efficacité à STC : η_{PV}^{STC}	0.195
Coefficient de température : TC	-0.0038
Température nominale opérationnelle : T^{NOCT} ($^{\circ}C$)	46

TABLEAU 4.B.1 : Spécifications techniques des panneaux photovoltaïques.

Le nombre de panneaux constituant le parc de production a été ici fixé à 3125, de sorte que la puissance nominale totale du parc est de 1 MW crête. De cette façon, tous les résultats obtenus sont normalisés et pourront donc être étendus à n'importe quel autre parc de production à travers un facteur de proportionnalité égal à la puissance (en MW crête) du nouveau parc de production.

Les prévisions probabilistes de la l'éclairement énergétique sont les prévisions d'ensemble ECMWF de l'éclairement énergétique global horizontal pour le site d'Adélaïde. La résolution spatiale de cette prévision étant de 0.2° , à la fois en latitude et longitude, la prévision pour le pixel le plus proche d'Adélaïde est prise comme prévision pour le site de production. Ces prévisions sont constituées d'un membre de contrôle (non perturbé) et de 50 membres perturbés.

4.B.2 Suréchantillonnage des prévisions

Les prévisions d'ensemble ECMWF (E_i^{GHI}) sont fournies avec un pas temporel de 3h. Il se trouve que pour la plupart des problématiques de gestion de l'électricité, une prévision avec une résolution temporelle maximale d'une heure est souhaitable. En effet, comme on le montre dans la partie 4.4.1, la valorisation de la production doit se faire au pas horaire pour des raisons réglementaires. Nous allons donc utiliser le suréchantillonnage (ou "oversampling"), qui permet d'augmenter la résolution temporelle d'une prévision. Les méthodes de suréchantillonnage sont très régulièrement utilisées dans les études liées à la météorologie (Dorado-Moreno et al., 2017).

Pour la plupart des applications météorologiques, un suréchantillonnage linéaire défini par l'équation 4.30 ne pose aucun problème particulier :

$$Y_{t_2} = Y_{t_1} + (t_2 - t_1) \times \frac{Y_{t_3} - Y_{t_1}}{t_3 - t_1}, \quad (4.30)$$

où Y_{t_2} est la variable suréchantillonnée à l'instant t_2 , situé entre t_1 et t_3 , de valeurs Y_{t_1} et Y_{t_3} connues.

Le cas de la variable "énergie" est cependant particulier, car on recherche souvent sa conservation. Or, elle n'est pas assurée par un suréchantillonnage linéaire. En revanche, elle est assurée lorsqu'on utilise une méthode de suréchantillonnage dite consistante. La propriété de consistance d'un suréchantillonnage peut s'exprimer ainsi : "la série temporelle suréchantillonnée puis re-moyennée au pas de temps de la série originale doit être égale à la série originale". Le suréchantillonnage linéaire a été appliqué sur des données réelles pour les journées du 03/01/2016 et du 04/01/2016. La figure 4.B.2 montre le problème de non-consistance.

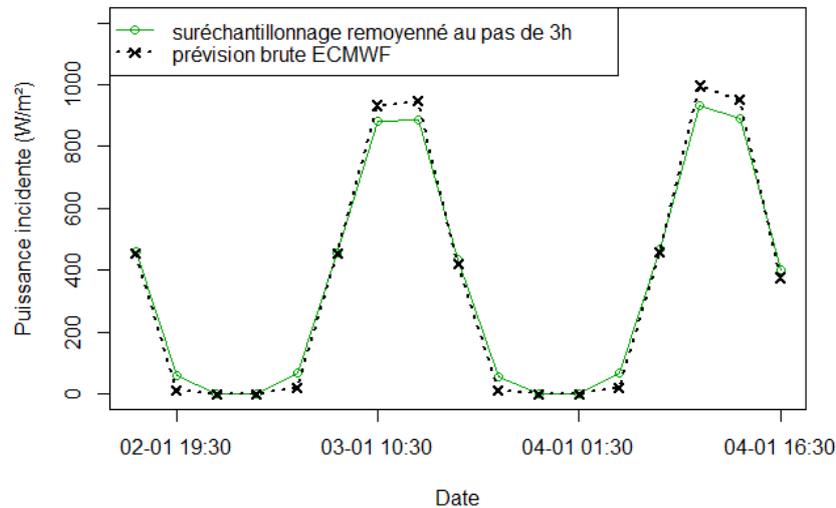


FIGURE 4.B.2 : Suréchantillonnage linéaire classique de l'éclairement énergétique.

Il est important de noter qu'un suréchantillonnage basé sur l'indice de ciel clair ne résout pas ce problème. C'est ce défaut identifié dans le suréchantillonnage linéaire qui a motivé l'utilisation d'une autre méthode. Nous avons choisi la méthode de suréchantillonnage linéaire avec correction de consistance proposée par le projet ENDORSE, détaillée par Espinar et al. (2011). Bien que d'autres méthodes consistantes ont déjà été proposées (Tikhonov et Arsenin, 1977 ; Rymes et Myers, 2001), la méthode sélectionnée a l'avantage d'avoir déjà été validée dans le cadre de l'éclairement énergétique. Cette méthode est définie comme étant la méthode consistante la plus proche, au sens de la norme L2, de la solution de suréchantillonnage "linéaire classique". C'est pourquoi la méthode choisie peut être appelée méthode de "suréchantillonnage linéaire avec correction de consistance". La figure 4.B.3 montre le suréchantillonnage obtenu à l'aide de cette méthode, et montre sa consistance sur les journées observées (les deux courbes se superposent), ce qui garantit la conservation de l'énergie entre la série temporelle originale et la série temporelle suréchantillonnée.

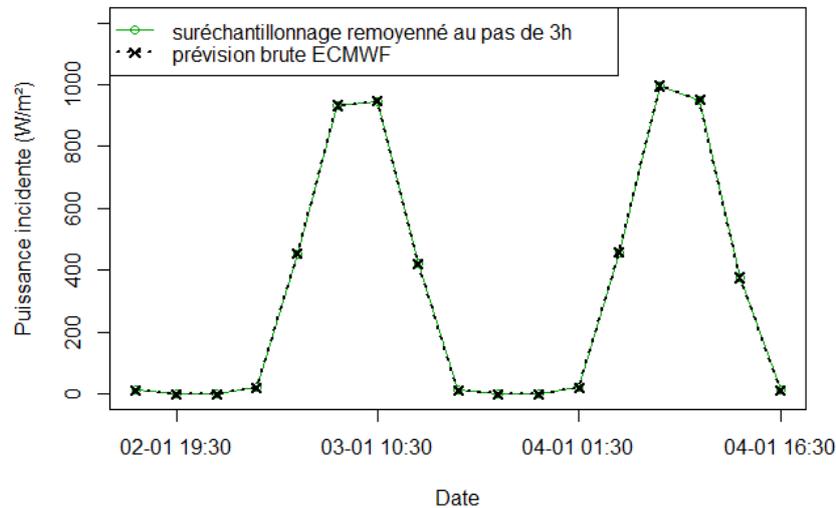


FIGURE 4.B.3 : Suréchantillonnage linéaire avec correction de consistance.

Cette méthode de suréchantillonnage a été appliquée à toutes les prévisions ECMWF au pas de 3h, afin d'obtenir une prévision suréchantillonnée (Eo_i^{GHI}) au pas horaire, sur toute la période considérée pour l'étude (soit les années 2016 et 2017).

4.B.3 Modélisation de la production électrique

Le modèle de production électrique est issu du livre de Luque et Hegedus (2011). Dans ce modèle, la température des modules T_{PV} dépend de la température ambiante T_{air} , de la température des modules observée sous les conditions de température nominale d'utilisation des cellules T^{NOCT} et de l'éclairement énergétique incident GHI selon :

$$T_{PV} = T_{air} + ((T^{NOCT} - 20)/800) \times GHI. \quad (4.31)$$

L'équation 4.32 permet le calcul de la production électrique $Prod_{PV}$, en prenant en compte la surface totale A , le coefficient d'efficacité η_{PV}^{STC} et le coefficient de température TC (cf. tableau 4.B.1) :

$$Prod_{PV} = A \times GHI \times \eta_{PV}^{STC} \times (1 - TC \times (T_{PV} - 25)). \quad (4.32)$$

Ce modèle est utilisé dans cette étude pour transformer une prévision d'ensemble suréchantillonnée de l'éclairement énergétique (Eo_i^{GHI}) en prévision d'ensemble de la production électrique (E_i^p), comme expliqué sur la figure 4.B.1. Ces ensembles sont donc liés par :

$$E_i^p = A \times Eo_i^{GHI} \times \eta_{PV}^{STC} \times (1 - TC \times (T_{PV} - 25)). \quad (4.33)$$

Enfin, des pertes diverses supplémentaires de 10 % ont été appliquées pour modéliser les pertes dues notamment aux onduleurs et à l'effet joule. Ce taux de 10 % a été choisi en concordance avec le taux de pertes diverses constatées par Daha (2017).

4.C Graphique de *quantile score* d'une prévision déterministe

Le graphique de *quantile score* d'une prévision déterministe \hat{F}_d est toujours une droite. Ceci est dû à la définition de la fonction de perte Pinball $\rho_{\tau=R}$ utilisée pour le calcul du *quantile score*, et dont le calcul dépend du signe de D (cf. chapitre 1). Cependant, pour ces 2 définitions, $\frac{d\rho}{d\tau}$ est constante.

Pour une prévision \hat{F}_τ et une observation O , on peut écrire :

$$\rho_\tau(\hat{F}_\tau, O) = \begin{cases} (O - \hat{F}_\tau) \times \tau & \text{si } O - \hat{F}_\tau \geq 0, \\ (O - \hat{F}_\tau) \times (\tau - 1) & \text{si } O - \hat{F}_\tau < 0. \end{cases} \quad (4.34)$$

Pour une prévision déterministe, $\hat{F}_\tau = \hat{F}_d$. Elle est donc caractérisée par le fait que \hat{F}_τ est constante et ne dépend pas de τ , ce qui n'est pas le cas pour une prévision probabiliste. Donc, dans le cas d'une prévision déterministe, on a simplement :

$$\frac{d\rho_\tau(\hat{F}_\tau, o)}{d\tau} = (O - \hat{F}_\tau). \quad (4.35)$$

Ainsi $QS(\tau)$ est une droite de pente $(O - \hat{F}_\tau)$. Pour un ensemble de couples prévision/observation, cette droite indique donc finalement le biais d'une prévision déterministe.

4.D Mise en place du modèle de prévision des prix ARIMA

Dans cette partie, la procédure de mise en place des modèles de prévision des prix est détaillée. L'objectif est de prévoir le ratio R et la somme γ pour chacun des trois pays. Pour simplifier le problème, le modèle de prévision a été optimisé pour la prévision du ratio R et le modèle déterminé ainsi est aussi appliqué pour la prévision de γ .

4.D.1 Étude des autocorrélations temporelles

Pour créer le modèle ARIMA de prévision des prix, la procédure générale recommandée par Contreras et al. (2003) a été suivie. Dans un premier temps, les autocorrélations des séries temporelles des différentes variables cibles sont étudiées sur les corrélogrammes des figures 4.D.1, 4.D.2, 4.D.3. Ces graphiques permettent de voir si un modèle ARIMA est adapté et donnent des indices sur les variables temporelles les plus importantes.

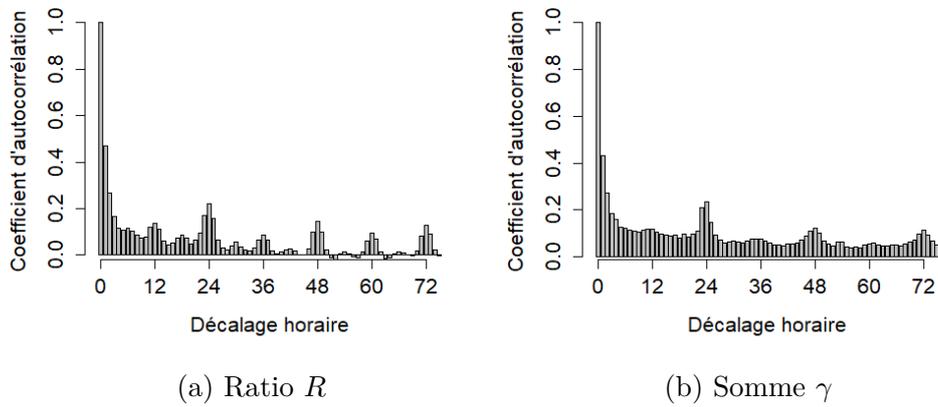


FIGURE 4.D.1 : Coefficients d'autocorrélation de (R, γ) pour la France.

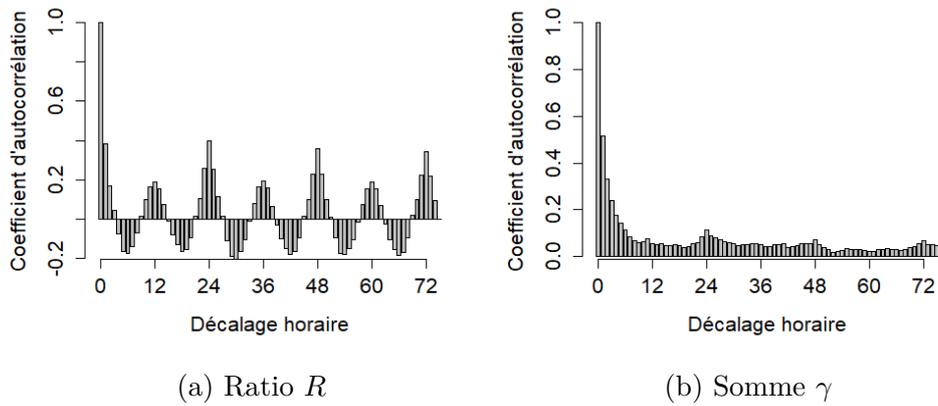


FIGURE 4.D.2 : Coefficients d'autocorrélation de (R, γ) pour le Portugal.

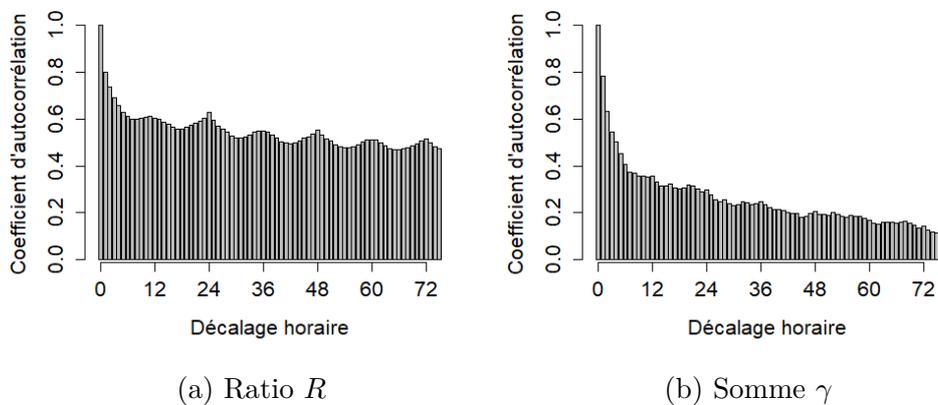


FIGURE 4.D.3 : Coefficients d'autocorrélation de (R, γ) pour la Suisse.

Les figures 4.D.1, 4.D.2 et 4.D.3 montrent des autocorrélations significatives, et il semble intéressant de les prendre en compte pour produire une prévision de qualité. Plus précisé-

ment, les autocorrélations dépendent du pays considérés : elles vont de fortes (en Suisse) à assez faibles (Portugal). En revanche, on constate pour toutes les variables de fortes périodicités : en règle général l'autocorrélation augmente légèrement toutes les 12 heures et fortement toutes les 24 heures. On constate même l'apparition d'autocorrélations négatives pour le ratio au Portugal. Elles ne pourraient s'expliquer que par la méthode de calcul des paramètres $\pi^{+,*}$ et $\pi^{-,*}$, ce qui n'est pas le sujet de cette étude.

4.D.2 Variables temporelles et ordres du modèle

Dans cette partie, on étudie les différentes variables temporelles qui pourraient avoir une influence sur les variables cibles. Trois variable saisonnières ont été identifiées : le mois de l'année, le jour de la semaine et l'heure de la journée. Pour déterminer si ces variables ont réellement une influence sur les variables de prévision, des tests ANOVA ("Analysis of variance") ont été réalisés. Les résultats de ces tests sur le ratio R sont présentés dans la table 4.D.1

Pays	Variable exogène		
	Heure de la journée	Jour de la semaine	Mois de l'année
France	2×10^{-16}	2.9×10^{-1}	1.6×10^{-4}
Portugal	2×10^{-16}	1.7×10^{-3}	2.5×10^{-4}
Suisse	2×10^{-16}	9.5×10^{-4}	2×10^{-16}

TABLEAU 4.D.1 : Indépendance (p-value) entre le ratio des pénalités de déséquilibre et les variables temporelles considérées.

Ces résultats ont mené à une sélection des variables utilisées pour la désaisonnalisation des variables cibles. Il n'existe pas, dans la littérature, de seuil absolu en dessous duquel une variable est jugée significative. Chaque prévision doit fixer son seuil, et il a été conseillé par Di Leo et Sardanelli (2020) de communiquer dans toute étude à la fois :

1. Les valeurs trouvées de p-value,
2. Le choix de la valeur-seuil de p-value en-dessous de laquelle une variable explicative est considérée significative.

Dans cette étude, nous avons choisi de prendre en compte dans la prévision seules les variables avec les plus faibles p-value (2×10^{-16}). La sélection des variables significatives est finalement résumée dans le tableau 4.D.2

Pays	Variable exogène		
	Heure de la journée	Jour de la semaine	Mois de l'année
France	✓	✗	✗
Portugal	✓	✗	✗
Suisse	✓	✗	✓

✓ Variable retenue ✗ Variable non retenue

TABLEAU 4.D.2 : Variables exogènes retenues pour la désaisonnalisation des variables cibles.

Comme il n'est pas possible d'utiliser les autocorrélations d'ordre inférieur à 12 à cause des contraintes opérationnelles (cf. partie 4.4.1), les périodicités les plus intéressantes à exploiter sont uniquement des multiples de 24 heures. Pour cette raison, un modèle ARIMA indépendant a été mis en place pour chaque heure de la journée. Pour le marché suisse, pour lequel le mois de l'année a été déterminé comme une variable à prendre en compte, les variables cibles ont été désaisonnalisées à l'aide de la méthode de décomposition "STL" (Cleveland et al., 1990).

Les ordres des méthodes ARIMA ont été déterminé grâce à la procédure proposée par Hyndman et Khandakar (2008). Par souci de simplicité, toutes les prévisions ont été faites avec les mêmes ordres pour un même pays. Finalement, les modèles ARIMA sont de la forme ARIMA(2,1,0) pour les marchés français et suisse et ARIMA(1,1,0) pour le marché portugais.

Conclusions générales et perspectives

Conclusions générales

Dans ce travail de thèse, nous avons proposé d'étudier les liens possibles entre qualité et valeur des prévisions solaires probabilistes. Pour y parvenir, une stratégie composée de deux étapes a été suivie. La première, qui se concentre sur la qualité des prévisions, est développée dans les trois premiers chapitres, alors que la deuxième s'attache à étudier la valeur et correspond au chapitre final. En parallèle de cette question de recherche principale, nous avons aussi souhaité démontrer que les prévisions probabilistes sont un outil qui pourrait être utilisé plus largement par les acteurs industriels. Pour cela, nous avons voulu évaluer les gains maximums de qualité et de valeur qu'un utilisateur peut espérer atteindre grâce à l'utilisation de prévisions probabilistes, construites à l'aide de prévisions d'ensemble générées par les modèles numériques de prévision du temps.

Il a été souligné que l'évaluation de la qualité doit s'appuyer sur des outils de vérification et des modèles de référence. Le chapitre 1 a mis l'accent sur l'importance d'utiliser des scores "propres", puis a recensé de nombreux outils de diagnostic de la qualité des prévisions probabilistes. Ce chapitre a aussi permis de présenter un nouvel outil graphique, appelé matrice d'espérance, créé dans le but de comparer entre eux les scores de prévisions probabilistes. Cet outil a montré que, parmi les scores présentés applicables aux prévisions de variables continues, les plus à même d'être utilisés pour mettre en évidence les relations entre qualité et valeur des prévisions sont le *CRPS* et le *quantile score*, car ils sont à la fois strictement propres et non divergents (donc plus à même de refléter la valeur d'une prévision que l'on conçoit difficilement infinie). Toujours à l'aide des matrices d'espérance, les gains potentiels de qualité de l'approche probabiliste évalués par le *CRPS* ont pu être quantifiés à environ 30 % dans le cas de prévisions non biaisées.

En vue de compléter le cadre d'évaluation de la qualité des prévisions solaires, nous avons proposé dans le chapitre 2 un nouveau modèle de référence climatologique appelé CSD-CLIM. Ce modèle est construit sur l'idée d'exploiter au mieux toute l'information contenue dans la seule climatologie. De cette façon, il est fiable par construction, et possède une qualité supérieure à la climatologie naïve et aux modèles de référence climatologiques proposés par la communauté scientifique comme le CH-PeEn.

Dans le chapitre 3, des modèles de prévision couramment utilisés dans la littérature scientifique ont été implémentés afin de produire des prévisions J+1 sur 6 sites présentant un large éventail de conditions climatologiques. Une nouvelle méthodologie appelée MC-CRPS a été introduite puis utilisée afin de déterminer sur un cas d'application pratique les gains effectifs en qualité de l'approche probabiliste par rapport à l'approche déterministe.

Ces gains effectifs varient entre 27.4 % et 34.6 % en terme de CRPS sur les différents sites considérés. De leur côté, les gains effectifs obtenus grâce à l'utilisation des prévisions d'ensemble pour la construction des prévisions probabilistes varient entre 4% et 16%.

Enfin, le chapitre 4 a été dédié à l'étude de la valeur des prévisions. Une recherche théorique sur la manière de modéliser au mieux les coûts subis par un utilisateur dans un contexte de prise de décision a été menée. Une prévision peut aider à optimiser cette prise de décision. Une structure de coût particulière appelée "market-based" a été mise en avant : elle correspond à celle que subit un acteur d'un marché de l'électricité dont les coûts dépendent entièrement des prix de ce marché. Dans ce cas, cette recherche a permis d'établir que le *quantile score* est un score probabiliste adapté pour retranscrire la valeur d'une prévision probabiliste. Cela a conduit à la proposition d'une méthodologie d'étude de la valeur des prévisions probabilistes de variables continues. Celle-ci recommande l'utilisation du graphique nommé EVC qui combine la représentation de la distribution des risques (associée au contexte des prix du marché) et celle de la performance de la prévision pour tous les niveaux de probabilités. Nous avons alors pu conclure que le gain maximal de l'approche probabiliste peut varier entre 0 et 100%, et dépend de la distribution des risques. Cette méthodologie a été utilisée pour expliquer les résultats obtenus sur un cas pratique de vente d'électricité sur trois marchés présentant des distributions de risque radicalement différentes. Les gains en valeur des approches probabiliste et ensembliste ont ensuite été quantifiés. Sur les trois marchés considérés, les gains de l'approche probabiliste par rapport à l'approche déterministe sont de 15%, 30% et 68%. Cependant, dans un contexte d'utilisation opérationnelle, il est nécessaire de disposer également d'une prévision des prix du marché. Celle-ci a aussi une importance significative sur les résultats obtenus et peut diminuer la valeur des prévisions probabilistes.

Certaines métriques spécifiques comme le *quantile score* ou le *CRPS*, créées à l'origine pour l'étude de la qualité, sont apparues naturellement dans la deuxième étape de ce travail traitant de la valeur. Finalement, ces outils constituent le lien entre ces deux notions. S'il peut exister des cas particuliers d'utilisation d'une prévision pour lesquels la valeur ne peut pas être formellement jugée grâce à ceux-ci, on peut néanmoins supposer que l'usage de ces métriques constitue toujours une première approximation pertinente. Tous les cas considérés dans cette étude suggèrent que l'utilisation de l'approche probabiliste permet d'augmenter sensiblement la valeur des prévisions.

Perspectives de travail

Le score du modèle CSD-CLIM, appelé *CSD-UNC*, peut être vu comme une caractérisation des conditions climatologiques d'un site. Cette interprétation semble prometteuse et ouvrir des pistes de recherche. La création d'une classification des sites en fonction des caractéristiques de leur courbe *CSD-UNC* pourrait être pertinente. Il est probable que l'on puisse distinguer des propriétés communes en fonction de la climatologie ou de la situation géographique du site.

Le travail effectué dans le chapitre 4 a mis en avant l'importance fondamentale des fonctions pertes. L'étude théorique menée a abouti à des résultats valables pour un type de fonctions particulier, représentant les coûts directs du marché. Les résultats obtenus

correspondent donc en première approximation aux coûts auxquels font face la plupart des acteurs industriels. Cependant, certains peuvent être exposés à des structures de coûts plus complexes (par exemple grâce à des flexibilités leur permettant d'être moins dépendants des prix du marché). Il semble alors nécessaire de travailler à une modélisation plus fidèle des coûts propres de tels acteurs, en utilisant des fonctions pertes différentes. Un premier pas a été effectué en ce sens dans ce travail, et a montré que la forme des fonctions pertes a bien une influence majeure sur les résultats obtenus. Tout cela considéré, il semblerait opportun de continuer à diversifier ces fonctions, afin de refléter plusieurs structures de coûts. Il conviendrait de travailler en partenariat avec des acteurs industriels dans le but de suggérer des structures adéquates et en phase avec leurs besoins.

Enfin, il a été mis en évidence l'influence majeure de la prévision des prix et de sa qualité sur le résultat final de l'optimisation. Afin de proposer un outil applicable au niveau industriel, il semble nécessaire de travailler plus longuement sur cette prévision. Par exemple, on pourrait proposer un modèle d'optimisation intégrant des prévisions probabilistes autant pour la production électrique que pour les prix.

Bibliographie

Abdallah, L. et El-Shennawy, T. (2013).

“Reducing Carbon Dioxide Emissions from Electricity Sector Using Smart Electric Grid Applications”. In : Journal of Engineering (cf. p. 3).

AEMO (2020).

Fact sheet: The national Electricity market. AEMO (cf. p. 141).

Ahmed, S. et al. (2020).

“Grid Integration Challenges of Wind Energy: A Review”. In : IEEE Access 8, p. 10857-10878 (cf. p. 6).

Aïd, R., Gruet, P. et Pham, H. (2016).

“An optimal trading problem in intraday electricity markets”.

In : Mathematics and Financial Economics 10 (1) (cf. p. 10).

Alessandrini, S., Sperati, S. et Pinson, P. (2013).

“A comparison between the ECMWF and COSMO Ensemble Prediction Systems applied to short-term wind power forecasting on real data”. In : Applied Energy 107, p. 271-280 (cf. p. 174).

Alessandrini, S. et al. (2015).

“An analog ensemble for short-term probabilistic solar power forecast”.

In : Applied Energy 157,

p. 95-110 (cf. p. 31, 59, 60, 62, 70, 88).

Amaral Burghi, A. do, Hirsch, T. et Pitz-Paal, R. (2020).

“Artificial Learning Dispatch Planning with Probabilistic Forecasts: Using Uncertainties as an Asset”. In : Energies 13.3, p. 616 (cf. p. 151).

Arrêté du 23 avril 2008 relatif aux prescriptions techniques de conception et de fonctionnement pour le raccordement à un réseau public de distribution d'électricité en basse tension ou en moyenne tension d'une installation de production d'énergie électrique (2008) (cf. p. 6).

Azimi, R., Ghayekhloo, M. et Ghofrani, M. (2016).

“A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting”.

In : Energy Conversion and Management 118,

p. 331-344 (cf. p. 17).

Bacher, P., Madsen, H. et Nielsen, H. A. (2009).

“Online short-term solar power forecasting”. en. In : Solar Energy 83.10,

p. 1772-1783 (cf. p. 88).

Badosa, J., Haeffelin, M. et Chepfer, H. (2013).

- “Scales of spatial and temporal variation of solar irradiance on Reunion tropical island”. In : Solar Energy 88,
p. 42-56 (cf. p. 100, 113).
- Bakker, K. et al. (2019).
“Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation”. In : Solar Energy 191,
p. 138-150 (cf. p. 88).
- Baran, S. et Lerch, S. (2018).
“Combining predictive distributions for the statistical post-processing of ensemble forecasts”. In : International Journal of Forecasting 34.3,
p. 477-496 (cf. p. 94).
- Behera, M. K., Majumder, I. et Nayak, N. (2018).
“Solar photovoltaic power forecasting using optimized modified extreme learning machine technique”. In : Engineering Science and Technology 21,
p. 428-438 (cf. p. 125).
- Ben Bouallegue, Z. (2017).
“Verification and post-processing of ensemble weather forecasts for renewable energy applications”. Thèse de doct. (cf. p. 21, 25, 136).
- Ben Bouallègue, Z., Pinson, P. et Friederichs, P. (2015).
“Quantile forecast discrimination ability and value”.
In : Quarterly Journal of the Royal Meteorological Society 141.693,
p. 3415-3424 (cf. p. 30).
- Benedetti, R. (2010).
“Scoring Rules for Forecast Verification”. In : Monthly Weather Review 138.1,
p. 203-211 (cf. p. 28).
- Bentzien, S. et Friederichs, P. (2014).
“Decomposition and graphical portrayal of the quantile score: Quantile Score Decomposition and Portrayal”.
In : Quarterly Journal of the Royal Meteorological Society 140.683,
p. 1924-1934 (cf. p. 29, 30).
- Bernardo, J. (1979).
“Expected Information as Expected Utility”. In : The Annals of Statistics 7.3,
p. 686-690 (cf. p. 28).
- Bickel, J. E. (2007).
“Some Comparisons among Quadratic, Spherical, and Logarithmic Scoring Rules”.
In : Decision Analysis 4.2,
p. 49-65 (cf. p. 27).
- Bilan Énergétique de la Réunion 2019 (2020). Observatoire Énergie Réunion (OER)
(cf. p. 4).
- Binter, R. (2012).
“Applied Probabilistic Forecasting”.
Thèse de doct. Department of Statistics of the London School of Economics
(cf. p. 60).
- Bird, R. E. et Hulstrom, R. L. (1981).
Simplified Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces.
Rapp. tech. Solar Energy Research Institute, Golden, CO (cf. p. 63, 113).
- Bourry, F. et al. (2008).

- “Advanced strategies for wind power trading in short-term electricity markets”.
 In : European Wind Energy Conference & Exhibition EWEC 2008,
 8 p. (Cf. p. 130).
- BP (2020).
Statistical Review of World Energy, 69ème édition. BP (cf. p. 2).
- Bremnes, J. B. (2004).
 “Probabilistic wind power forecasts using local quantile regression”.
 In : Wind Energy 7.1,
 p. 47-54 (cf. p. 132).
- Bridier, L., David, M. et Lauret, P. (2014).
 “Optimal design of a storage system coupled with intermittent renewables”.
 In : Renewable Energy 67,
 p. 2-9 (cf. p. 7).
- Brier, G. (1950).
 “Verification of forecasts in terms of probability”. In : Monthly Weather Review 78.1
 (cf. p. 18, 25).
- Bröcker, J. (2009a).
Decompositions of Proper Scores. Rapp. tech.
 Max-Planck-Institut für Physik komplexer Systeme (cf. p. 19, 28).
- (2009b).
 “Reliability, Sufficiency, and the Decomposition of Proper Scores”.
 In : Quarterly Journal of the Royal Meteorological Society 135.643,
 p. 1512-1519 (cf. p. 19, 24).
- Bröcker, J. et Smith, L. A. (2007a).
 “Scoring Probabilistic Forecasts: The Importance of Being Proper”.
 In : Weather and Forecasting 22,
 p. 382-388 (cf. p. 23, 25, 27).
- Bröcker, J. et Smith, L. A. (2007b).
 “Increasing the Reliability of Reliability Diagrams”.
 In : Weather and Forecasting 22.3,
 p. 651-661 (cf. p. 33, 34, 83).
- Bucksteeg, M., Niesen, L. et Weber, C. (2016).
 “Impacts of Dynamic Probabilistic Reserve Sizing Techniques on Reserve
 Requirements and System Costs”.
 In : IEEE Transactions on Sustainable Energy 7.4,
 p. 1408-1420 (cf. p. 10).
- Buizza, R. (2008).
 “The value of probabilistic prediction”. In : Atmospheric Science Letters 9.2,
 p. 36-42 (cf. p. 59, 88, 127).
- Candille, G. et Talagrand, O. (2005).
 “Evaluation of probabilistic prediction systems for a scalar variable”.
 In : Quarterly Journal of the Royal Meteorological Society 131,
 p. 2131-2150 (cf. p. 24, 25, 32, 65, 72).
- Casati, B. et al. (2008).
 “Forecast verification: current status and future directions”.
 In : Meteorological Applications 15,
 p. 3-18 (cf. p. 17).

- Chen, S., Gooi, H. et Wang, M. (2013).
 “Solar radiation forecast based on fuzzy logic and neural networks”.
 In : Renewable Energy 60,
 p. 195-201 (cf. p. 17).
- Chernozhukov, V., Fernández-Val, I. et Galicho, A. (2010).
 “Quantile and Probability Curves Without Crossing”. In : Econometrica 78 (3),
 p. 1093-1125 (cf. p. 92).
- Christensen, H. M. (2015).
 “Decomposition of a New Proper Score for Verification of Ensemble Forecasts”.
 In : Monthly Weather Review 143.5,
 p. 1517-1532 (cf. p. 30).
- Cleveland, R. B. et al. (1990).
 “STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with
 Discussion)”. In : Journal of Official Statistics 6,
 p. 3-73 (cf. p. 183).
- Cludius, J. et al. (2014).
 “The merit order effect of wind and photovoltaic electricity generation in Germany
 2008–2016: Estimation and distributional implications”. In : Energy Economics 44,
 p. 302-313 (cf. p. 145).
- Coimbra, C. F., Kleissl, J. et Marquez, R. (2013).
 “Overview of Solar-Forecasting Methods and a Metric for Accuracy Evaluation”. en.
 In : Solar Energy Forecasting and Resource Assessment. Elsevier,
 p. 171-194 (cf. p. 97).
- Commission Européenne (2018).
 “Europe leads the global clean energy transition: Commission welcomes ambitious
 agreement on further renewable energy development in the EU”.
 In : <https://ec.europa.eu> (cf. p. 3).
- Contreras, J. et al. (2003).
 “ARIMA models to predict next-day electricity prices”.
 In : IEEE Transactions on Power Systems 18.3,
 p. 1014-1020 (cf. p. 153, 180).
- Cros, S. et al. (2013).
 “Clear sky models assessment for an operational PV production forecasting solution”.
 In : 28th European Photovoltaic Solar Energy Conference and Exhibition (cf. p. 67).
- Cusack, S. et Arribas, A. (2008).
 “Assessing the Usefulness of Probabilistic Forecasts”.
 In : Monthly Weather Review 136.4,
 p. 1492-1504 (cf. p. 60).
- Daha, H. D. (2017).
 “Modélisation et analyse expérimentale d’une centrale solaire photovoltaïque en
 milieu désertique maritime”. Thèse de doct. INSA Lyon (cf. p. 179).
- David, M. et al. (2016).
 “Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH
 models”. en. In : Solar Energy 133,
 p. 55-72 (cf. p. 59, 62, 63, 70, 100).
- David, M., Mazon Aguiar, L. et Lauret, P. (2018).

- “Comparison of intraday probabilistic forecasting of solar irradiance using only endogenous data”. en. In : International Journal of Forecasting 34.3, p. 529-547 (cf. p. 63, 100).
- Dawid, A. P. (1984).
 “Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach.”
 In : Journal of the Royal Statistical Society 147.2, p. 278-292 (cf. p. 34).
- (2006).
 “The geometry of proper scoring rules”. en. In : p. 17 (cf. p. 25, 56).
- Dawid, P., Lauritzen, S. et Parry, M. (2012).
 “Proper local scoring rules on discrete sample spaces”.
 In : The Annals of Statistics 40.1, p. 593-608 (cf. p. 25).
- de Sisternes, F. J., Jenkins, J. D. et Botterud, A. (2016).
 “The value of energy storage in decarbonizing the electricity sector”.
 In : Applied Energy 175, p. 368-379 (cf. p. 7).
- Declercq, C. et Vincent, A. (2000).
 “L’ouverture du marché de l’électricité”.
 In : Courrier hebdomadaire du CRISP 1684.19, p. 1-53 (cf. p. 142).
- Décret no 2017-530 du 12 avril 2017 relatif à la programmation pluriannuelle de l’énergie de La Réunion (2017) (cf. p. 4).
- Delle Monache, L. et al. (2013).
 “Probabilistic Weather Prediction with an Analog Ensemble”.
 In : Monthly Weather Review 141, p. 3498-3516 (cf. p. 92).
- Di Leo, G. et Sardanelli, F. (2020).
 “Statistical significance: p value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach”.
 In : European Radiology Experimental 4.1 (cf. p. 182).
- Diagne, H. M. (2015).
 “Intelligent management of electrical grid from La Reunion. Solar irradiance forecasting in an insular grid”. Theses. Université de la Réunion (cf. p. 3, 8).
- Dorado-Moreno, M. et al. (2017).
 “Combining Reservoir Computing and Over-Sampling for Ordinal Wind Power Ramp Prediction”. In : Advances in Computational Intelligence. Springer International Publishing, p. 708-719 (cf. p. 177).
- Doubleday, K., Van Scyoc Hernandez, V. et Hodge, B. (2020).
 “Benchmark probabilistic solar forecasts: Characteristics and recommendations”.
 In : Solar Energy 206, p. 52-67 (cf. p. 59-62, 65, 71, 98, 121).
- Ernst, D. (2016-2017).
Energy Markets (cf. p. 144).
- ERSE (2020).

- Manual do Procedimentos da Gestão Global do Sistema. ERSE (cf. p. 153).
- Escrig, H. et al. (2013).
 “Cloud detection, classification and motion estimation using geostationary satellite imagery for cloud cover forecast”. In : Energy 55,
 p. 853-859 (cf. p. 8).
- Espinar, B. et al. (2011).
Report on the harmonization and qualification of meteorological data. Rapp. tech. Armines (cf. p. 178).
- Forrest, D. et Simmons, R. (2000).
 “Forecasting sport: the behaviour and performance of football tipsters”.
 In : International Journal of Forecasting 16.3,
 p. 317-331 (cf. p. 11).
- Fortin, V. et al. (2014).
 “Why should ensemble spread match the RMSE of the ensemble mean?”
 In : Journal of Hydrometeorology 15,
 p. 1708-1713 (cf. p. 108).
- Giummolè, F. et Mameli, V. (2018).
 “Asymptotic minimum scoring rule prediction”.
 In : Electronic Journal of Statistics 12.2,
 p. 2401-2429 (cf. p. 27).
- Gneiting, T., Balabdaoui, F. et Raftery, A. (2007).
 “Probabilistic forecasts, calibration and sharpness.”
 In : Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69.2,
 p. 243-268 (cf. p. 21, 36, 47).
- Gneiting, T. et Raftery, A. E. (2007).
 “Strictly Proper Scoring Rules, Prediction, and Estimation”.
 In : Journal of the American Statistical Association 102,
 p. 359-378 (cf. p. 19, 27, 37).
- Gneiting, T. et al. (2005).
 “Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation”. In : Monthly Weather Review 133.5,
 p. 1098-1118 (cf. p. 64, 83, 89, 93).
- Golestaneh, F., Gooi, H. B. et Pinson, P. (2016a).
 “Generation and evaluation of space–time trajectories of photovoltaic power”.
 In : Applied Energy 13.3,
 p. 80-91 (cf. p. 151).
- Golestaneh, F., Pinson, P. et Gooi, H. (2016b).
 “Very Short-Term Nonparametric Probabilistic Forecasting of Renewable Energy Generation— With Application to Solar Energy”.
 In : IEEE Transactions on Power Systems 31.5,
 p. 3850-3863 (cf. p. 62).
- Hall, P. J. et Bain, E. J. (2008).
 “Energy-storage technologies and electricity generation”. In : Energy Policy 36.12.
 Foresight Sustainable Energy Management and the Built Environment Project,
 p. 4352-4355 (cf. p. 7).
- Hamill, T. et Whitaker, J. (2006).

- “Probabilistic Quantitative Precipitation Forecasts Based on Reforecast Analogs: Theory and Application”. In : Monthly Weather Review 134.11, p. 3209-3229 (cf. p. 89, 92).
- Hamill, T. M. et Colucci, S. (1997).
 “Verification of Eta–RSM Short-Range Ensemble Forecasts”.
 In : Monthly Weather Review 125.6, p. 1312-1327 (cf. p. 64).
- Hamill, T. M. (2001).
 “Interpretation of Rank Histograms for Verifying Ensemble Forecasts”.
 In : Weather and Forecasting 129.3, p. 550-560 (cf. p. 32, 33).
- Hartmann, H. et al. (2002).
 “Confidence Builders: Evaluating Seasonal Climate Forecasts from User Perspectives”. In : Bulletin of the American Meteorological Society 83.5, p. 683-698 (cf. p. 34).
- Hersbach, H. (2000).
 “Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems”. en. In : Weather and Forecasting 15.5, p. 559-570 (cf. p. 26, 65, 97, 99, 108).
- Hoff, T. E. et Perez, R. (2012).
 “Modeling PV fleet output variability”. en. In : Solar Energy 86.8, p. 2177-2189 (cf. p. 100).
- Hoff, T. E. et al. (2013).
 “Reporting of irradiance modeling relative prediction errors”.
 In : Progress in Photovoltaics: Research and Applications 21.7, p. 1514-1519 (cf. p. 95).
- Holton, J. (2012).
An Introduction to dynamic meteorology, 5th edition. Academic Press, p. 552 (cf. p. 9).
- Holttinen, H. (2005).
 “Optimal electricity market for wind power”. In : Energy Policy 33.16, p. 2052-2063 (cf. p. 130).
- Hopson, T. M. (2014).
 “Assessing the Ensemble Spread–Error Relationship”.
 In : Monthly Weather Review 142.3, p. 1125-1142 (cf. p. 108).
- Hyndman, R. J. et Khandakar, Y. (2008).
 “Automatic Time Series Forecasting: The forecast Package for R”.
 In : Journal of Statistical Software, Articles 27.3, p. 1-22 (cf. p. 183).
- Ineichen, P. et Perez, R. (2002).
 “A new airmass independent formulation for the Linke turbidity coefficient”. en.
 In : Solar Energy 73.3, p. 151-157 (cf. p. 63).
- Iversen, E. B. et al. (2014).

- “Probabilistic Forecasts of Solar Irradiance by Stochastic Differential Equations”.
In : Environmetrics 25,
p. 152-164 (cf. p. 88).
- Jakaša, T., Andročec, I. et Sprčić, P. (2011).
“Electricity price forecasting — ARIMA model approach”.
In : 2011 8th International Conference on the European Energy Market (EEM),
p. 222-225 (cf. p. 153).
- Jolliffe, I. T. et Stephenson, D. B. (2003).
Forecast Verification: A Practitioner’s Guide in Atmospheric Science. Wiley
(cf. p. 96).
- Jordan, G. et Piwko, R. (2011).
“The Value of Wind Power Forecasting”.
In : 91st American Meteorological Society Annual Meeting. Washington, DC
(cf. p. 125).
- Junk, C., Delle Monache, L. et Alessandrini, S. (2015).
“Analog-Based Ensemble Model Output Statistics”.
In : Monthly Weather Review 143.7,
p. 2909-2917 (cf. p. 89).
- Kalecinski, N. (2015).
“Processus de formation et d’étalement des nuages sur l’Ile de la Réunion:
caractérisation à partir de données issues d’observations satellite, sol et du modèle
numérique de prévision AROME; application à la prévision des énergies solaires.”
Thèse de doct. Ecole Polytechnique (cf. p. 100).
- Key World Energy Statistics (2017). Agence internationale de l’énergie (cf. p. 2).
- Koenker, R. et D’Orey, V. (1994).
“Confidence Intervals for Regression Quantiles”.
In : Journal of the Royal Statistical Society 36.3,
p. 383-393 (cf. p. 91).
- Kumler, A., Xie, Y. et Zhang, Y. (2019).
“A Physics-based Smart Persistence model for Intra-hour forecasting of solar
radiation (PSPI) using GHI measurements and a cloud retrieval technique”.
In : Solar Energy 177,
p. 494-500 (cf. p. 17).
- Laio, F. et Tamea, S. (2007).
“Verification tools for probabilistic forecasts of continuous hydrological variables”.
In : Hydrology and Earth System Sciences Discussions 11.4,
p. 1267-1277 (cf. p. 30).
- Lauret, P., David, M. et Pedro, H. (2017).
“Probabilistic Solar Forecasting Using Quantile Regression Models”.
In : Energies 10.10,
p. 1591 (cf. p. 20, 62, 88).
- Lauret, P., David, M. et Pinson, P. (2019).
“Verification of solar irradiance probabilistic forecasts”. In : Solar Energy 194,
p. 254-271 (cf. p. 20, 59, 64, 65, 89, 90, 95-98, 101, 134, 174).
- Lauret, P., Lorenz, E. et David, M. (2016).
“Solar Forecasting in a Challenging Insular Context”. In : Atmosphere 7.2 (cf. p. 7).
- Lauret, P. et al. (2015).

- “A benchmarking of machine learning techniques for solar radiation forecasting in an insular context”. In : Solar Energy 112, p. 446-457 (cf. p. 7).
- Lave, M., Kleissl, J. et Stein, J. S. a. (2013).
 “A Wavelet-Based Variability Model (WVM) for Solar PV Power Plants”.
 In : IEEE Transactions on Sustainable Energy 4.2, p. 501-509 (cf. p. 4).
- Le Gal La Salle, J. et al. (2020).
 “Added-value of Ensemble Prediction System on the quality of solar irradiance probabilistic forecasts”. In : Renewable Energy (cf. p. 63, 64, 74, 85).
- Lefèvre, M. et al. (2013).
 “McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions”. In : Atmos. Meas. Tech., p. 17 (cf. p. 63).
- Lerch, S. et Thorarinsdottir T., L. (2013).
 “Comparison of non-homogeneous regression models for probabilistic wind speed forecasting”. In : Tellus A: Dynamic Meteorology and Oceanography 65.1, p. 21206 (cf. p. 89, 94).
- Leutbecher, M. et Palmer, T. N. (2008).
 “Ensemble forecasting”. In : Journal of Computational Physics 227.7, p. 3515-3539 (cf. p. 9, 88, 90).
- Li, M. (2017).
World Energy 2017-2050: Annual Report (cf. p. 3).
- Li, T. Y. et Yorke, J. A. (1975).
 “Period Three Implies Chaos”. In : The American Mathematical Monthly 10, p. 985-992 (cf. p. 11).
- Linnet, U. (2005).
 “Tools supporting wind energy trade in deregulated markets”.
 Thèse de doct. Technical University of Denmark (cf. p. 132, 133).
- Loi n° 2015-992 du 17 août 2015 relative à la transition énergétique pour la croissance verte (1) (2015) (cf. p. 7).
- Long, C. N. et Dutton, E. G. (2010).
BSRN Global Network recommended QC tests, V 2.0. Rapp. tech. PANGAEA (cf. p. 113).
- Long, C. N. et Shi, Y. (2008).
 “An Automated Quality Assessment and Control Algorithm for Surface Radiation Measurements”. In : The Open Atmospheric Science Journal 2.1, p. 23-37 (cf. p. 26, 66).
- Lorenz, E. (1963).
 “Deterministic Nonperiodic Flow”. In : Journal of the Atmospheric Sciences 20.2, p. 130-141 (cf. p. 10).
- Lorenz, E. et al. (2009a).
 “Benchmarking of different approaches to forecast solar irradiance”. In : (cf. p. 7).
- Lorenz, E. et al. (2009b).
 “Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems”. In :
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

- 2.1,
p. 2-10 (cf. p. 88).
- Luque, A. et Hegedus, S. (2011).
Handbook of photovoltaic science and engineering.
Wiley, Chichester, 195 West Sussex, U.K. (cf. p. 179).
- Machete, R. L. (2013).
“Contrasting Probabilistic Scoring Rules”.
In : Journal of statistical Planning and Inference 143,
p. 1781-1790 (cf. p. 26).
- Massidda, L. et Marrocu, M. (2018).
“Quantile Regression Post-Processing of Weather Forecast for Short-Term Solar
Power Probabilistic Forecasting”. In : Energies 11.7,
p. 1763 (cf. p. 88).
- Matevosyan, J. et Soder, L. (2006).
“Minimization of imbalance cost trading wind power on the short-term power
market”. In : IEEE Power & Energy Magazine 21,
p. 1396-1404 (cf. p. 150).
- Murphy, A. H. (1969).
“On the “Ranked Probability Score””. In : Journal of Applied Meteorology 8.6,
p. 988-989 (cf. p. 25).
- (1973).
“A new vector partition of the probability score”.
In : Journal of Applied Meteorology 12.4,
p. 595-600 (cf. p. 60).
- Murphy, A. H. (1966).
“A Note on the Utility of Probabilistic Predictions and the Probability Score in the
Cost-Loss Ratio Decision Situation”.
In : Journal of Applied Meteorology and Climatology 5.4,
p. 534-537 (cf. p. 127).
- (1993).
“What Is a Good Forecast? An Essay on the Nature of Goodness in Weather
Forecasting”. In : Weather and Forecasting 8.2,
p. 281-293 (cf. p. 11, 17, 20, 125).
- Orwig, K. et al. (2012).
“Enhanced Short-Term Wind Power Forecasting and Value to Grid Operations:
Preprint”. In : (cf. p. 10).
- Pedro, H. et al. (2018).
“Assessment of machine learning techniques for deterministic and probabilistic
intra-hour solar forecasts”. In : Renewable Energy 123,
p. 191-203 (cf. p. 62, 97).
- Pelland, S. et al. (2013).
Photovoltaic and Solar Forecasting: State of the Art. Rapp. tech. IEA PVPS
(cf. p. 125).
- Perez, R. et al. (2019).
“Operationally Perfect Solar Power Forecasts: A Scalable Strategy to Lowest-Cost
Firm Solar Power Generation”.

- In : 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC). T. 2, p. 1-6 (cf. p. 12).
- Perez, R. et al. (2007).
 “Forecasting solar radiation – Preliminary evaluation of an approach based upon the national forecast database”. In : Solar Energy 81.6, p. 809-812 (cf. p. 17).
- Perez, R. et al. (2020).
 “From Firm Solar Power Forecasts to Firm Solar Power Generation an Effective Path to Ultra-High Renewable Penetration a New York Case Study”.
 In : Energies 13.17 (cf. p. 7).
- Pierro, M. et al. (2019).
 “Residual load probabilistic forecast for reserve assessment: a real case study”.
 In : Renewable Energy 149, p. 99-110 (cf. p. 10, 59, 87).
- Pinson, P. (2012).
 “Adaptive calibration of (u,v)-wind ensemble forecasts”.
 In : Quarterly Journal of the Royal Meteorological Society 138.666, p. 1273-1284 (cf. p. 89).
- Pinson, P., Chevallier, C. et Kariniotakis, G. (2004).
 “Optimizing Benefits from wind power participation in electricity market using advanced tools for wind power forecasting and uncertainty assessment”.
 In : EWEC 2004. Londres, United Kingdom (cf. p. 132).
- (2007).
 “Trading Wind Generation From Short-Term Probabilistic Forecasts of Wind Power”. In : IEEE Transactions on Power Systems 22.3, p. 1148-1156 (cf. p. 131, 133, 149).
- Pinson, P., Juban, J. et Kariniotakis, G. N. (2006).
 “On the Quality and Value of Probabilistic Forecasts of Wind Generation”. In : 2006 International Conference on Probabilistic Methods Applied to Power Systems, p. 1-7 (cf. p. 35).
- Pinson, P. et Madsen, H. (2009).
 “Ensemble-based probabilistic forecasting at Horns Rev”. In : Wind Energy 12.2, p. 137-155 (cf. p. 89).
- Pinson, P. et Tastu, J. (2014).
 “Discussion of “Prediction Intervals for Short-Term Wind Farm Generation Forecasts” and “Combined Nonparametric Prediction Intervals for Wind Power Generation””. In : IEEE Transactions on Sustainable Energy 5.3, p. 1019-1020 (cf. p. 25).
- Pinson, P., McSharry, P. et Madsen, H. (2010).
 “Reliability diagrams for non-parametric density forecasts of continuous variables: Accounting for serial correlation”. en.
 In : Quarterly Journal of the Royal Meteorological Society 136.646, p. 77-90 (cf. p. 33, 34).
- Ramahatana, F. (2021).
 “Intégration des prévisions probabilistes dans le contrôle des systèmes énergétiques : application à un microréseau.” unpublished, to be submitted.
 Thèse de doct. Université de la Réunion (cf. p. 150).

- Ramahatana, F. et David, M. (2019).
 “Economic optimization of micro-grid operations by dynamic programming with real energy forecast”. In : J. Phys.: Conf. Ser. 10 (cf. p. 10).
- Richardson, D. S. (2000).
 “Skill and relative economic value of the ECMWF ensemble prediction system”.
 In : Quarterly Journal of the Royal Meteorological Society 126.563,
 p. 649-667 (cf. p. 127).
- Roebber, P. (2009).
 “Visualizing Multiple Measures of Forecast Quality”.
 In : Weather and Forecasting 24.2,
 p. 601-608 (cf. p. 17).
- Roulston, M. S. et Smith, L. A. (2002).
 “Evaluating Probabilistic Forecasts Using Information Theory”.
 In : Monthly Weather Review 130,
 p. 1653-1660 (cf. p. 28).
- Roulston, M. S. et al. (2003).
 “Using medium-range weather forecasts to improve the value of wind energy production”. In : Renewable Energy 28.4,
 p. 585-602 (cf. p. 11, 12, 125).
- RTE (2020).
Règles MA-RE. RTE (cf. p. 153).
- Rymes, M. D. et Myers, D. R. (2001).
 “Mean preserving algorithm for smoothly interpolating averaged data”.
 In : Solar Energy 71.4,
 p. 225-231 (cf. p. 178).
- Scheuerer, M. (2014).
 “Probabilistic quantitative precipitation forecasting using Ensemble Model Output Statistics: Probabilistic Precipitation Forecasting Using EMOS”.
 In : Quarterly Journal of the Royal Meteorological Society 140.680,
 p. 1086-1096 (cf. p. 94).
- Schlömer, S. et al. (2014).
 “Annex III: Technology-specific cost and performance parameters. In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Edenhofer, O., R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C. Minx (eds.)]” In : (cf. p. 3).
- Schneider, M. et al. (2008).
 “Qualified Forecast of Ensemble Power Production by Spatially Dispersed Grid-Connected PV Systems”. In : (cf. p. 7).
- Schueller, G. I., Gumbel, E. et Panggabean, H. (1976).
 “Probabilistic determination of design wind velocity in Germany”.
 In : Proceedings of the Institution of Civil Engineers 61.4,
 p. 673-683 (cf. p. 11).
- Shannon, C. E. (1948).

- “A mathematical theory of communication”.
In : The Bell System Technical Journal 27.3,
p. 379-423 (cf. p. 28).
- Shaomao, L. (2017).
“Trading Strategies and Risk Management for Wind Power in the Electricity
Market”. Thèse de doct. Auburn University (cf. p. 149, 150).
- Shen, B. et al. (2021).
“Is Weather Chaotic?: Coexistence of Chaos and Order within a Generalized Lorenz
Model”. In : Bulletin of the American Meteorological Society 102.1,
E148-E158 (cf. p. 11).
- Sims, C. (1993).
“A Nine-Variable Probabilistic Macroeconomic Forecasting Model”.
In : Business Cycles, Indicators, and Forecasting.
National Bureau of Economic Research, Inc,
p. 179-212 (cf. p. 11).
- Sloughter, J. M., Gneiting, T. et Raftery, A. (2010).
“Probabilistic Wind Speed Forecasting Using Ensembles and Bayesian Model
Averaging”. In : Journal of the American Statistical Association 105.489,
p. 25-35 (cf. p. 11).
- Sperati, S., Alessandrini, S. et Delle Monache, L. (2016).
“An application of the ECMWF Ensemble Prediction System for short-term solar
power forecasting”. In : Solar Energy 133,
p. 437-450 (cf. p. 88, 174).
- Stephenson, D. B., Coelho, C. A. S. et Jolliffe, I. T. (2008).
“Two Extra Components in the Brier Score Decomposition”.
In : Weather and Forecasting 23.4,
p. 752-757 (cf. p. 55).
- Strähl, C. et Ziegel, J. (2017).
“Cross-calibration of probabilistic forecasts”.
In : Electronic Journal of Statistics 11.1,
p. 608-639 (cf. p. 21).
- Swissgrid (2020).
Prescriptions générales de groupe-bilan. Swissgrid (cf. p. 153).
- Talagrand, O., Vautard, R. et Strauss, B. (1997).
“Evaluation of probabilistic prediction systems”.
In : Workshop on Predictability, 20-22 October 1997.
ECMWF. Shinfield Park, Reading : ECMWF,
p. 1-26 (cf. p. 32).
- Tapachès, E. et al. (2019).
“Mitigation of the Variability of a PV Fleet via Geographical Dispersion and Energy
Storage Systems on the Reunion Island Non-Interconnected Grid”.
In : 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC),
p. 2082-2086 (cf. p. 7).
- Thorey, J., Chaussin, C. et Mallet, V. (2018).
“Ensemble forecast of photovoltaic power with online CRPS learning”.
In : International Journal of Forecasting 34.4,
p. 762-773 (cf. p. 62).

- Thornes, J. E. et Stephenson, D. B. (2001).
 “How to judge the quality and value of weather forecast products”.
 In : Meteorological Applications 8.3,
 p. 307-314 (cf. p. 17, 126).
- Tikhonov, A. N. et Arsenin, V. Y. (1977).
Solutions of Ill-Posed Problems. New York : Winston (cf. p. 178).
- Todter, J. (2011).
 “New Aspects of Information Theory in Probabilistic Forecast Verification”.
 Thèse de doct. Goethe University (cf. p. 28).
- Todter, J. et Ahrens, B. (2012).
 “Generalization of the Ignorance Score: Continuous Ranked Version and Its
 Decomposition”. In : Monthly Weather Review 140.6,
 p. 2005-2017 (cf. p. 28, 66, 68).
- Tsyplakov, A. (2013).
 “Evaluation of Probabilistic Forecasts: Proper Scoring Rules and Moments”.
 In : SSRN Electronic Journal (cf. p. 21, 22).
- Vallance, L. (2018).
 “Synergie des mesures pyranométriques et des images hémisphériques in-situ avec
 des images satellites météorologiques pour la prévision photovoltaïque”.
 Thèse de doct. Université Paris sciences et lettres (cf. p. 5).
- Vallance, L. et al. (2017).
 “Towards a standardized procedure to assess solar forecast accuracy: A new ramp
 and time alignment metric”. In : Solar Energy 150,
 p. 408-422 (cf. p. 17).
- Van der Meer, D. et al. (2020).
 “Clear-sky index space-time trajectories from probabilistic solar forecasts:
 Comparing promising copulas”.
 In : Journal of Renewable and Sustainable Energy 12.2 (cf. p. 151).
- Van der Meer, D., Widén, J. et Munkhammar, J. (2018).
 “Review on probabilistic forecasting of photovoltaic power production and electricity
 consumption”. In : Renewable and Sustainable Energy Reviews 81,
 p. 1484-1512 (cf. p. 59).
- Vannitsem, S., Wilks, D. et Messner, J. (2018a).
Statistical Postprocessing of Ensemble Forecasts. Elsevier (cf. p. 64).
- (2018b).
Statistical Postprocessing of Ensemble Forecasts. Elsevier (cf. p. 89).
- Wang, Q. et al. (2016).
 “The value of improved wind power forecasting: Grid flexibility quantification, ramp
 capability analysis, and impacts of electricity market operation timescales”.
 In : Applied Energy 184,
 p. 696-713 (cf. p. 125).
- Weigel, A. P. et Mason, S. J. (2011).
 “The Generalized Discrimination Score for Ensemble Forecasts”.
 In : Monthly Weather Review 139.9,
 p. 3069-3074 (cf. p. 20).
- Weijs, S., Van Nooijen, R. et Van de Giesen, N. (2010).

- “Kullback–Leibler Divergence as a Forecast Skill Score with Classic Reliability–Resolution–Uncertainty Decomposition”.
In : Monthly Weather Review 138.9,
p. 3387-3399 (cf. p. 28).
- Weron, R. (2014).
“Electricity price forecasting: A review of the state-of-the-art with a look into the future”. In : International Journal of Forecasting 30.4,
p. 1030-1081 (cf. p. 153).
- West, S. et al. (2014).
“Short-term irradiance forecasting using skycams: Motivation and development”.
In : Solar Energy 110,
p. 188-207 (cf. p. 9).
- What is an ensemble forecast? (s. d.). Metoffice (cf. p. 9).
- Whitaker, J. S. et Lough, A. F. (1998).
“The Relationship between Ensemble Spread and Ensemble Mean Skill”.
In : Monthly Weather Review 126.12,
p. 3292-3302 (cf. p. 108).
- Wilks, D. S. (2001).
“A skill score based on economic value for probability forecasts”.
In : Meteorological Applications 8.2,
p. 209-219 (cf. p. 126).
- Wilks, D. (2014).
Statistical Methods in the Atmospheric Sciences. Academic Press
(cf. p. 20, 41, 96, 97, 127).
- Wilks, D. S. (2006).
“Comparison of ensemble-MOS methods in the Lorenz ’96 setting”.
In : Meteorological Applications 13.3,
p. 243 (cf. p. 89).
- Williams, R. M., Ferro, C. A. T. et Kwasniok, F. (2014).
“A comparison of ensemble post-processing methods for extreme events”.
In : Quarterly Journal of the Royal Meteorological Society 140.680,
p. 1112-1120 (cf. p. 89).
- Winkler, R. L. et al. (1996).
“Scoring rules and the evaluation of probabilities”. In : Test 5.1,
p. 1-60 (cf. p. 29).
- Winkler, R. L. (1994).
“Evaluating Probabilities: Asymmetric Scoring Rules”. en.
In : Management Science 40.11,
p. 1395-1405 (cf. p. 29).
- World Energy Outlook (2020). Agence internationale de l’énergie (cf. p. 2).
- World Radiation Monitoring Center-Baseline Surface Radiation Network (s. d.).
<https://bsrn.awi.de/>. Accessed: 2021-02-01. Alfred-Wegener-Institute
(cf. p. 14, 63).
- Woschnagg, E. et Cipan, J. (2019).
Evaluating forecast accuracy (cf. p. 154).
- Copernicus Atmosphere Monitoring Service (s. d.).

- CAMS McClear Clear-Sky Irradiation Service, version 3.1.
<http://www.soda-pro.com/fr/web-services/radiation/cams-mcclear>.
Accessed: 2021-02-01 (cf. p. 63).
- Yang, D. (2019).
“A universal benchmarking method for probabilistic solar irradiance forecasting”.
In : Solar Energy 184,
p. 410-416 (cf. p. 59-61, 63, 66, 70, 71, 79).
- Yang, D. (2020).
“Choice of clear-sky model in solar forecasting”.
In : Journal of Renewable and Sustainable Energy 12.2,
p. 026101 (cf. p. 63).
- Yang, D., Van der Meer, D. et Munkhammar, J. (2020a).
“Probabilistic solar forecasting benchmarks on a standardized dataset at Folsom,
California”. In : Solar Energy 206,
p. 628-639 (cf. p. 59).
- Yang, D. et al. (2020b).
“Verification of deterministic solar forecasts”.
In : Solar Energy 210. Special Issue on Grid Integration,
p. 20-37 (cf. p. 17, 60).
- Yates, J. (1982).
“External correspondence: Decompositions of the mean probability score”.
In : Organizational Behavior and Human Performance 30.1,
p. 132-156 (cf. p. 55).
- Yitzhaki, S. (2003).
“Gini’s Mean difference: a superior measure of variability for non-normal
distributions”. In : Metron - International Journal of Statistics 61,
p. 285-316 (cf. p. 94).
- Young, R. (2010).
“Decomposition of the Brier score for weighted forecast-verification pairs: Weighted
Brier Score Decomposition”.
In : Quarterly Journal of the Royal Meteorological Society 136.650,
p. 1364-1370 (cf. p. 25).
- Yuen, R. et al. (2018).
ensembleMOS: Ensemble Model Output Statistics. R package version 0.8.2
(cf. p. 94).
- Zamo, M. et al. (2014a).
“A benchmark of statistical regression methods for short-term forecasting of
photovoltaic electricity production, part I: Deterministic forecast of hourly
production”. In : Solar Energy 105,
p. 792-803 (cf. p. 7).
- Zamo, M. et al. (2014b).
“A benchmark of statistical regression methods for short-term forecasting of
photovoltaic electricity production. Part II: Probabilistic forecast of daily
production”. In : Solar Energy 105,
p. 804-816 (cf. p. 88).
- Zhu, Y. et al. (2002).

“The economic value of ensemble-based weather forecasts”.

In : Bulletin of the American Meteorological Society 83.1,

p. 73-83 (cf. p. 59, 88, 126).

Zwaenepoel, B. et al. (2014).

“Solar commercial virtual power plant day ahead trading”.

In : 2014 IEEE PES General Meeting | Conference Exposition,

p. 1-5 (cf. p. 10).

Qualité et valeur des prévisions solaires probabilistes

Résumé

La prévision solaire est un élément central de la plupart des solutions d'intégration à large échelle de l'énergie solaire au sein des mix électriques. L'objectif de ce travail est de démontrer la valeur ajoutée des prévisions solaires probabilistes, encore peu utilisées à l'échelle industrielle. Pour cela, un cadre d'évaluation de la qualité des prévisions probabilistes est présenté. Il permet de quantifier les gains potentiels de qualité que l'utilisation de prévisions probabilistes permet d'espérer, puis de mesurer dans un second temps les gains réellement obtenus pour plusieurs modèles de prévision.

La valeur des prévisions du point de vue de l'utilisateur est mesurée sur des exemples simples, puis la question des liens entre qualité et valeur est explorée. Le rôle prépondérant de la structure des coûts s'appliquant aux décisions prises à l'aide de la prévision apparaît clairement. De celle-ci dépend fortement la valeur effective des prévisions probabilistes.

Mots-clés

Prévision solaire, Prévision numérique du temps, Prévision probabiliste, Valeur des prévisions, Scores propres

Quality and value of solar probabilistic forecasting

Abstract

Solar forecasting is a major asset for the large-scale penetration of solar energy into electricity mixes. This work aims at proving the added-value of solar probabilistic forecasting, not yet frequently used by market players. A framework for quality assessment of probabilistic forecasting is presented, and helps quantifying the potential relative gains of the probabilistic approach. Afterwards, the effective benefits of probabilistic predictions are measured on case studies.

The value of forecasts from the user perspective is measured on simple study cases, and the relationship between quality and value is investigated. The significance of the cost structure underlying the decision-making process associated to the forecast is highlighted. The effective value of probabilistic forecast strongly depends on this cost structure.

Keywords

Solar forecasting, Numerical Weather Prediction, Probabilistic forecasting, Value of forecasts, Proper scores