



# Smart meters, open data et IA

**Vers la réalisation d'un jumeau digital du Swiss Energypark** | Chaque jour, près d'un million de données sont récoltées dans le laboratoire en conditions réelles qu'est le Swiss Energypark. Afin de pouvoir les exploiter au mieux et les mettre à disposition de divers projets de recherche, le Swiss Energypark a initié la réalisation d'un jumeau digital. Une étape se fait en collaboration avec la Haute école Arc.

LAURENT RAEBER, HATEM GHORBEL, ALEXIS MAIRE, FARID ABDALLA

**L**e Swiss Energypark est une région pilote (voir encadré) qui a pour but de contribuer à la transition énergétique en accueillant des projets de différentes natures, mais poursuivant le même objectif: augmenter l'autoconsommation à l'échelle d'un réseau. Certains consistent à installer un dispositif physique dans le réseau (projets « hardware »), d'autres sont liés à un échange d'informations (projet « software »), quant aux derniers, ils étudient et accompagnent l'évolution de nos comportements (projets comportementaux).

Pour répondre aux besoins de ces divers projets, le Swiss Energypark collecte depuis plus de dix ans près d'un

million de données par jour. Cette profusion de mesures permet aux projets d'évaluer leur potentiel soit par comparaison avec des données historiques, soit par simulation dans un jumeau digital du système. Il est toutefois nécessaire, dans un cas comme dans l'autre, de rassembler, d'organiser et de formater toutes ces données de natures et de sources différentes afin de pouvoir les exploiter.

## Une véritable mine de données

Des mesures couvrant plusieurs années, qui comprennent des données de consommation extrêmement fines et un mix énergétique renouvelable varié, représentent une valeur inesti-

mable pour de très nombreux projets de recherche. Que ce soit pour simuler l'impact d'un décalage temporel de l'utilisation de certains appareils sur le réseau ou l'utilisation de batteries pour le stockage des surplus de production, ou encore pour évaluer une évolution des comportements au fil du temps: les exemples pour lesquels avoir accès à ce type de données constitue un élément essentiel sont innombrables. En effet, comme l'a si bien dit Lord Kelvin: « Si tu ne peux pas le mesurer, tu ne peux pas l'améliorer ». Et c'est précisément pour pouvoir mettre ces données à disposition que le Swiss Energypark a initié la construction d'un double digital.

## Vers la réalisation d'un jumeau digital

Le Swiss Energypark a défini trois phases dans l'intégration des données nécessaires à la création d'un double digital :

- l'intégration des données de production et de consommation à l'échelle du réseau ;
- l'intégration des données des smart meters (compteurs intelligents) ;
- et la modélisation du réseau physique lui-même.

Au cours de la première phase, un « data lake » (lac de données) a été créé pour stocker les données de production et de consommation à l'échelle du réseau. Les résultats sont visibles en temps réel sur le site du Swiss Energypark [2].

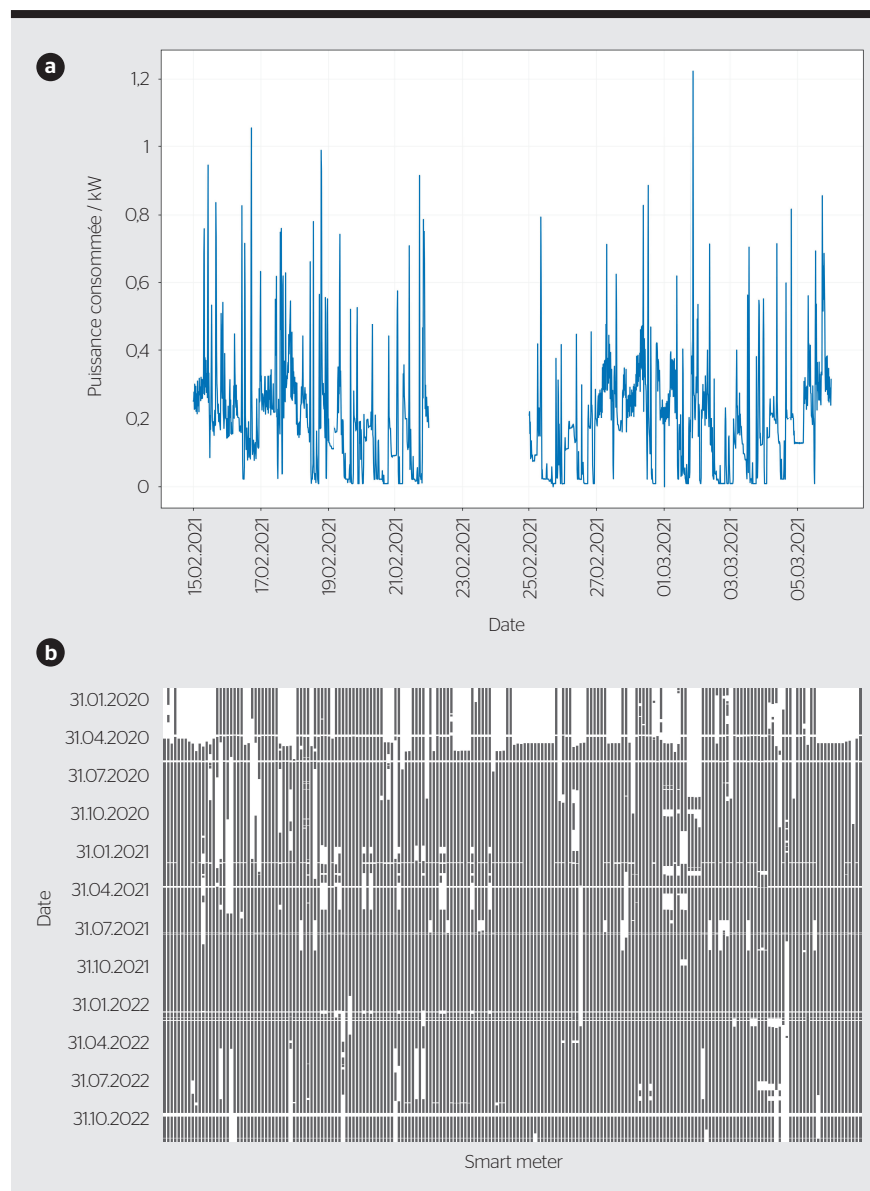
L'intégration des données de consommation issues des smart meters (deuxième phase) présente également un intérêt particulier, car ces dernières sont abondantes et couvrent plusieurs années. Les deux premières phases de l'élaboration de ce jumeau digital permettront notamment de mesurer en temps réel l'adéquation entre, d'une part, une production décentralisée et variée d'électricité renouvelable et, d'autre part, la consommation et les tendances sociétales qui se dessinent.

Une autre application consiste à extraire des lots de données pertinents pour les besoins d'un projet de recherche spécifique, et ce, afin de simuler et/ou de comparer l'effet d'un élément ajouté ou l'évolution d'un comportement. La complexité majeure de ce type de développement nécessite cependant des compétences importantes en matière de digitalisation, de data science et d'intelligence artificielle (IA). C'est là que l'écosystème du Swiss Energypark entre en action.

## Une collaboration avec trois objectifs

Le Swiss Energypark et la Haute école Arc (HE-Arc) ont débuté une collaboration à l'automne 2022 en vue d'atteindre divers objectifs.

Il s'agit, dans un premier temps, de **contrôler la plausibilité des données et de compléter les données manquantes**. En effet, si seules quelques mesures annuelles de consommation sont nécessaires au gestionnaire de réseau pour assurer le suivi requis et adéquat, ce n'est pas forcément le cas pour certains projets de recherche qui



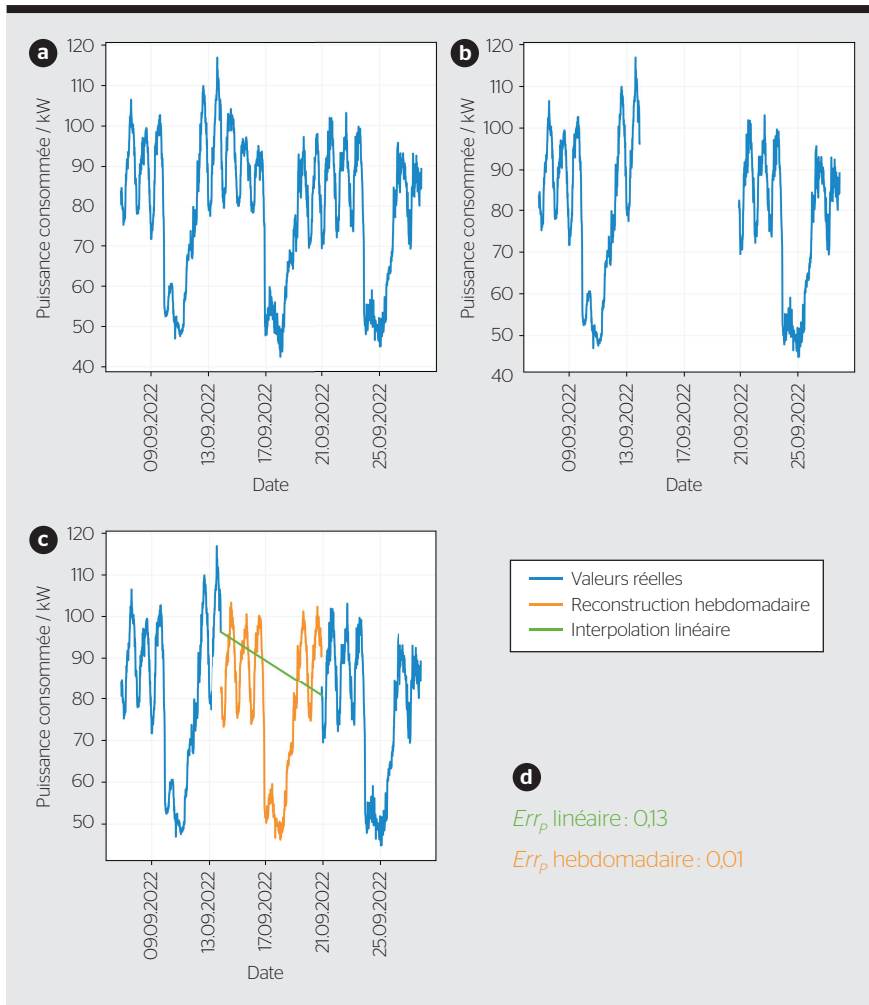
**Figure 1 a)** Exemple de profil de charge comportant des données manquantes. **b)** Synthèse des données manquantes sur un échantillon de trois ans de mesures et de 200 smart meters : les rectangles blancs correspondent aux plages de données manquantes.

nécessitent des données aussi complètes que possible. Il est donc nécessaire de reconstruire d'éventuelles données manquantes parmi les milliers de mesures supplémentaires effectuées afin d'avoir une série temporelle complète, puis d'évaluer l'incertitude induite par ce processus d'extrapolation.

Le deuxième objectif consiste à **extraire des sets de données d'intérêt**. Il s'agit là d'être à même de fournir des jeux de données pertinents en fonction de différents paramètres liés à un projet spécifique. En effet, chaque projet a ses caractéristiques et ses objectifs ; l'un souhaite étudier la consom-

mation de certains ménages, l'autre une certaine région ou encore un type de bâtiments présentant certaines caractéristiques. Extraire des données répondant à des critères est primordial et permet à un projet de se concentrer sur ses buts grâce à des données plausibles, triées et organisées.

Enfin, il faut encore pouvoir **fournir les données sous forme ouverte (open data)**. Il s'agit ici de mettre à disposition l'historique des données de consommation disponibles du Swiss Energypark en « open access », pour qu'elles puissent être utilisées facilement dans le cadre de projets de recherche. Les mesures de consom-



**Figure 2** Évaluation de deux procédés de reconstruction de données : **a**) série initiale de données réelles, **b**) création artificielle d'un trou de données, **c**) reconstruction des données manquantes via une approche hebdomadaire (en orange) et via une approche naïve (interpolation linéaire, en vert), et **d**) évaluation des résultats via le calcul de l'erreur de puissance relative.

tion couvrant plusieurs années sont en effet rares à cette échelle. Grâce à ces dernières, il est possible, par exemple, de mesurer des évolutions, de déceler des similitudes de consommation, ou encore de calculer des journées types représentatives permettant une première approche avec une quantité de données limitée et fiable.

Dans un premier temps, la HE-Arc a concentré ses efforts sur l'atteinte du premier de ces objectifs, à savoir sur la reconstruction des données manquantes dans un set de données.

### État des lieux des données de consommation

Le Swiss Energypark dispose d'environ 8000 compteurs intelligents (smart meters) qui enregistrent les consommations électriques des utilisateurs du

réseau (au pas de temps horaire ou toutes les 15 minutes, en fonction des modèles), et ce, depuis plus de 10 ans.

Une première analyse des données brutes enregistrées sur la période 2018-2023 par l'ensemble des smart meters a permis de mettre en évidence un certain nombre de données manquantes dans ces séries temporelles de consommation électrique. La **figure 1** illustre cette problématique avec, notamment, une synthèse visuelle des données manquantes pour un échantillon de 200 consommateurs entre 2020 et 2023. Ces plages sont assez fréquentes et peuvent durer de quelques heures à quelques mois. Il est toutefois problématique de travailler avec ces «trous» temporels. C'est pourquoi une méthodologie de reconstruction des données manquantes a été développée.

### Reconstruction des données manquantes

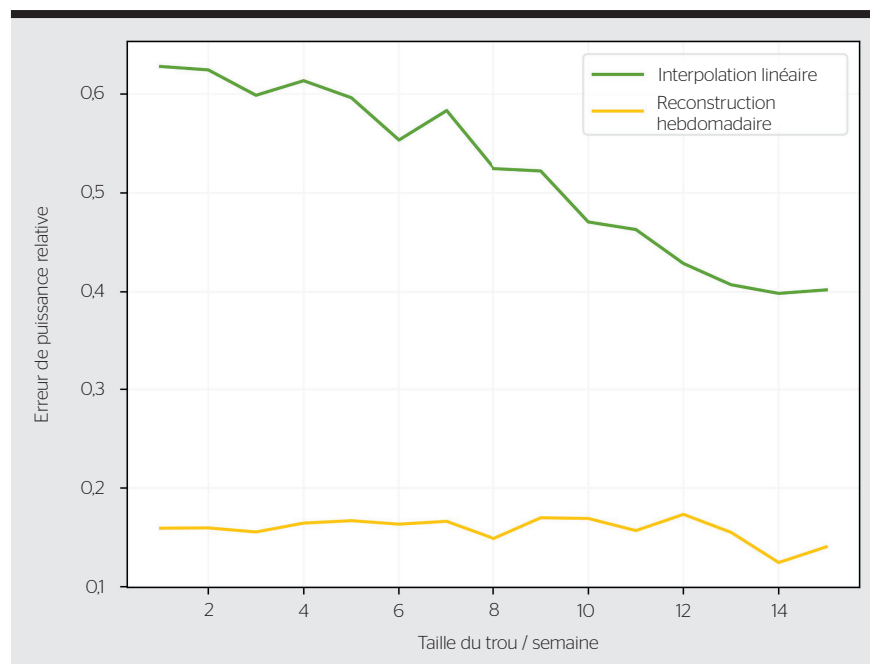
Dans le cadre des séries temporelles, et notamment dans le domaine de l'énergie, les méthodes de reconstruction ou de prédiction de valeurs sont nombreuses et diverses [3]. La reconstruction des données mise en place repose sur l'hypothèse simple que les courbes de charge dessinent un motif hebdomadaire. À un instant  $t$ , si la valeur de consommation est manquante dans la série, il suffit d'aller chercher les valeurs de consommation présentes au même instant la semaine précédente et la semaine suivante, de faire la moyenne de ces deux valeurs, puis d'utiliser cette dernière pour remplacer la valeur manquante. Si les valeurs sont également manquantes une semaine plus tôt et/ou une semaine plus tard, il suffit d'aller chercher deux semaines avant et deux semaines après, et ainsi de suite.

## Le Swiss Energypark en quelques mots...

Le Swiss Energypark regroupe une production électrique éolienne avec, notamment, les turbines de la centrale éolienne de Mont-Crosin exploitée par la société Juvent SA, une production hydroélectrique composée principalement du barrage de la Goule, et une production photovoltaïque qui comprend, entre autres, la centrale solaire de Mont-Soleil. L'ensemble de ces productions est injecté dans le réseau des Forces Électriques de la Goule. Ces installations - exclusivement renouvelables - permettent de couvrir entre 75 et 95 % des besoins annuels des 21 000 habitants approvisionnés par le réseau de la Goule. De plus, l'installation de smart meters dès les années 2010 permet de connaître la consommation pratiquement en temps réel. La variété des sources d'énergies renouvelables, le taux d'autonomie annuel et la connaissance précise de la consommation au cours de ces dix dernières années font du Swiss Energypark un «laboratoire en conditions réelles» unique en Suisse [1].

## Évaluation du procédé de reconstruction

Évaluer le procédé mis en place requiert de pouvoir comparer les valeurs reconstruites aux valeurs réelles, qui sont inconnues dans le cas des plages de données manquantes. Des trous artificiels de différentes durées (variant de 1 à 15 semaines) ont donc été créés parmi les données réelles dans le but de reconstruire ces valeurs manquantes et de pouvoir les comparer aux vraies valeurs (figure 2). Différents indicateurs sont couramment utilisés dans la littérature pour évaluer la prédiction de séries temporelles [4], et il faut tenir compte ici du fait que les courbes de charge peuvent contenir des valeurs d'ordres de grandeur très différents en fonction du consommateur (foyer privé ou grande entreprise). La puissance électrique totale reconstruite sur une plage de données manquantes a ensuite été comparée à la puissance électrique réelle en calculant l'«erreur de puissance relative». Cette dernière permet de mesurer la performance du procédé de reconstruction à l'échelle de la longueur de la plage de données manquantes, mais ne caractérise pas la reconstruction à la granularité de la série temporelle.



**Figure 3** Évolution de l'erreur de puissance relative en fonction de la durée de la plage de données manquantes pour les cas de la reconstruction hebdomadaire mise en place et d'une simple interpolation linéaire. Chaque valeur correspond à la moyenne de 1000 expériences tirées au sort parmi le panel de données de consommation disponibles.

Pour évaluer le procédé de reconstruction, une approche naïve a été également testée à titre de comparaison. Celle-ci consiste en une interpo-

lation linéaire des valeurs manquantes à partir de la dernière valeur disponible avant la plage et de la première disponible après la plage.

### IN KÜRZE

## Smart Meter, Open Data und künstliche Intelligenz

Realisierung eines digitalen Zwillings des Swiss Energypark

Der Swiss Energypark ist ein in der Schweiz einzigartiges «Labor unter realen Bedingungen». Er vereint insbesondere die Windkraft des Mont-Crosin, die Wasserkraftanlagen von La Goule und das Sonnenkraftwerk Mont-Soleil, die alle in ein einziges Stromnetz einspeisen. Ausserdem sammelt er mit Hilfe von Smart Metern, die seit den 2010er-Jahren installiert wurden, täglich fast eine Million Verbrauchsdaten. Diese Daten sind eine wahre Goldgrube für diverse Forschungsprojekte im Energiesektor. Um sie für solche Projekte zur Verfügung stellen zu können, hat der Swiss Energypark die Realisierung eines digitalen Zwillings initiiert.

Es wurden drei Phasen identifiziert, die für die Schaffung dieses digitalen Doppelgängers nötig sind: die Integration der netzweiten Produktions- und Verbrauchsdaten, die Integration der Smart-Meter-Daten und schliesslich die Modellierung des Stromnetzes. Die hohe Komplexität dieser Art von Entwicklung erfordert jedoch umfangreiche Kompetenzen in den Bereichen Digitalisierung, Data Science und künstliche Intelligenz (KI). Um die Integration

der Verbrauchsdaten zu realisieren, haben der Swiss Energypark und die Haute école Arc (HE-Arc) im Herbst 2022 eine Zusammenarbeit begonnen, um verschiedene Ziele zu erreichen: die Plausibilität der gesammelten Daten zu überprüfen und fehlende Daten zu ergänzen, Datensätze zu extrahieren, die für verschiedene Projekte von Interesse sind, und schliesslich die Daten in offener Form (open data) zur Verfügung zu stellen.

Zunächst konzentriert die HE-Arc ihre Arbeit auf die Rekonstruktion fehlender Daten in Datensätzen sowie auf die Bewertung und Verbesserung des Rekonstruktionsverfahrens. Parallel dazu hat der Swiss Energypark das Ziel, bald eine erste Stichprobe anonymisierter Verbrauchsdaten als Open Data zur Verfügung zu stellen. Der Rest wird folgen, sobald ein für die Daten des Swiss Energypark geeigneter Datenanonymisierungsdienst durch eine Partnerschaft mit dem Schweizer Zentrum für Künstliche Intelligenz (Centre Suisse d'Intelligence Artificielle, CSIA) entwickelt worden ist.

Cette expérience a été répétée 1000 fois, et les résultats ont été agrégés à partir de l'ensemble des expériences pour offrir des mesures plus robustes. Quelle que soit la durée de la plage de données manquantes, la puissance totale consommée sur la plage est reconstruite en moyenne avec un écart d'environ 15% par rapport à la puissance totale réelle: il s'agit là de résultats sensiblement meilleurs que ceux obtenus avec la méthode naïve, qui aboutit à un écart de 40% dans le meilleur des cas (figure 3).

### Vers une reconstruction des données plus précise

La reconstruction des données manquantes mise en place devrait pouvoir être améliorée par la mise à disposition prochaine des index relatifs à la facturation des consommateurs du Swiss Energypark. Ces index seront utilisés pour normaliser la reconstruction des plages de données manquantes de telle sorte que la puissance totale consommée reconstruite reflète bien la réalité et que l'erreur de puissance relative tende vers zéro.

### Prochains développements

Le Swiss Energypark a pour objectif de mettre à disposition très prochaine-

ment un premier échantillon de données de consommation anonymisées en open data pour qu'elles puissent être utilisées par des projets de recherche, et ce, avant de fournir la totalité des données dans un second temps. Les informations liées à l'accès et aux données disponibles seront mises à jour sur le site du Swiss Energypark [2].

Pour ce faire, le Swiss Energypark et la HE-Arc ont souhaité s'engager dans un partenariat à plus long terme. Le Swiss Energypark a donc choisi de rejoindre, en tant que partenaire, le Centre Suisse d'Intelligence Artificielle pour les PME (CSIA-PME) [5], porté par la Haute école spécialisée de Suisse occidentale (HES-SO). Cette nouvelle structure, regroupant cinq équipes de la HES-SO – dont l'équipe Analyse de données de la HE-Arc –, vise à démocratiser l'intelligence artificielle au sein des PME suisses en leur mettant à disposition des outils et des services leur permettant d'accélérer le déploiement de composants IA au sein de leur entreprise.

Les objectifs principaux de ce partenariat consisteront à développer un service d'anonymisation des données adapté aux données du Swiss Energypark, et ce, afin de permettre de publier l'ensemble des données de

consommation en open data. Les équipes de la HE-Arc auront également pour tâche de mettre à disposition des jeux de données sur mesure pour divers projets de recherche des partenaires du Swiss Energypark.

#### Références

- [1] Valérie Bourdin, « Une région pilote de la transition énergétique », Bulletin SEV/VSE 6/2022, p. 14-16, 2022. [bulletin.ch/fr/news-detail/une-region-pilote-de-la-transition-energetique.html](http://bulletin.ch/fr/news-detail/une-region-pilote-de-la-transition-energetique.html)
- [2] [swiss-energypark.com](http://swiss-energypark.com)
- [3] Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, Kwok Wei Shah, « A review on time series forecasting techniques for building energy consumption », Renewable and Sustainable Energy Reviews, Vol. 74, p. 902-924, 2017. [doi.org/10.1016/j.rser.2017.02.085](https://doi.org/10.1016/j.rser.2017.02.085).
- [4] Kadir Amasyali, Nora M. El-Gohary, « A review of data-driven building energy consumption prediction studies », Renewable and Sustainable Energy Reviews, Vol. 81, Part 1, p. 1192-1205, 2018. [doi.org/10.1016/j.rser.2017.04.095](https://doi.org/10.1016/j.rser.2017.04.095).
- [5] [swiss-ai-center.ch/fr/accueil](http://swiss-ai-center.ch/fr/accueil)

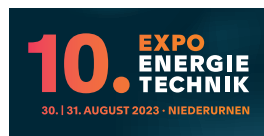
#### Auteurs

**Laurent Raeber** est directeur du Swiss Energypark et de la Société Mont-Soleil.  
→ [swiss-energypark.com](mailto:swiss-energypark.com)  
→ [laurent.raeber@swiss-energypark.com](mailto:laurent.raeber@swiss-energypark.com)

**Hatem Ghorbel** est professeur et responsable du groupe Analyse de données de la HE-Arc.  
→ HE-Arc, 2000 Neuchâtel  
→ [hatem.ghorbel@he-arc.ch](mailto:hatem.ghorbel@he-arc.ch)

**Alexis Maire** est adjoint scientifique au sein du groupe Analyse de données de la HE-Arc.  
→ [alexis.maire@he-arc.ch](mailto:alexis.maire@he-arc.ch)

**Farid Abdalla** est assistant scientifique au sein du groupe Analyse de données de la HE-Arc.  
→ [farid.abdalla@he-arc.ch](mailto:farid.abdalla@he-arc.ch)



TOGA Trafoklemmen  
und Abdeckhauben



**pronutec AG**

Starkstromkomponenten  
von den Experten



Composants basse tension  
par des experts

pronutec AG  
Rosenweg 3  
6234 Triengen

2017-2022  
**5**  
Jahre  
années

041 545 86 70  
[info@pronutec.ch](mailto:info@pronutec.ch)  
[www.pronutec.ch](http://www.pronutec.ch)